PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification: C07K 14/00 — A2

(11) International Publication Number: WO 01/04148

(43) International Publication Date: 18 January 2001 (18.01.2001)

(21) International Application Number: PCT/US00/19066

(22) International Filing Date: 13 July 2000 (13.07.2000)

(30) Priority Data:
60/143,582 13 July 1999 (13.07.1999) US
60/144,359 16 July 1999 (16.07.1999) US
60/184,442 23 February 2000 (23.02.2000) US

(60) Parent Application or Grant
MEDIMMUNE, INC. [/]; (). HULTGREN, Scott, J. [/]; (). PINKNER, Jerome, S. [/]; (). SAUER, Frederic [/]; (). BARNHART, Michelle [/]; (). WAKSMAN, Gabriel [/]; (). KNIGHT, Stefan [/]; (). GRANT, Alan, J.; ().

Publish d

(54) Title: DONOR STRAND COMPLEMENTED PILIN AND ADHESIN BROAD-BASED VACCINES

(54) Titre: VACCINS DE PILINE ET D'ADHESINE COMPLETES PAR UN BRIN DONNEUR POUR PRODUCTION A GRANDE ECHELLE

(57) Abstract

Modified polypeptides containing a pilus-protein sequence and a donor complementary sequence are disclosed, as well as complexes formed of a modified polypeptide and a pilin, or pilus-protein. Also disclosed are methods of using these novel polypeptides as a means of preventing and/or treating bacterial induced diseases, especially those caused by enterobacteria such as Escherichia coli. Methods of employing these modified pilus-derived polypeptides and complexes as vaccines and for generating antibodies for further study as well as for clinical purposes are also disclosed herein. In addition, processes for large scale production of antibacterial vaccines containing said polypeptides and complexes are also described.

(57) Abrégé

L'invention concerne des polypeptides modifiés contenant une séquence de protéine de pilus et une séquence complémentaire de donneur, ainsi que des complexes constitués d'un polypeptide modifié et d'une piline, ou protéine de pilus. L'invention concerne également des procédés d'utilisation de ces nouveaux polypeptides comme moyen de prévention et/ou de traitement de maladies provoquées par des bactéries, spécialement celles dues à des entérobactéries telles que Escherichia coli. L'invention concerne aussi des procédés d'utilisation de ces polypeptides et complexes dérivés de pilus comme vaccins et pour produire des anticorps en vue d'études ultérieures et à des fins cliniques. De plus, l'invention concerne des procédés de production à grande échelle de vaccins antibactériens contenant lesdits polypeptides et complexes.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 January 2001 (18.01.2001)

PCT

(10) International Publication Number
WO 01/04148 A2

(51) International Patent Classification[7]: C07K 14/00

(21) International Application Number: PCT/US00/19066

(22) International Filing Date: 13 July 2000 (13.07.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/143,582	13 July 1999 (13.07.1999)	US
60/144,359	16 July 1999 (16.07.1999)	US
60/184,442	23 February 2000 (23.02.2000)	US

(71) Applicant: MEDIMMUNE, INC. [US/US]; 35 West Watkins Mill Road, Gaithersburg, MD 20878 (US).

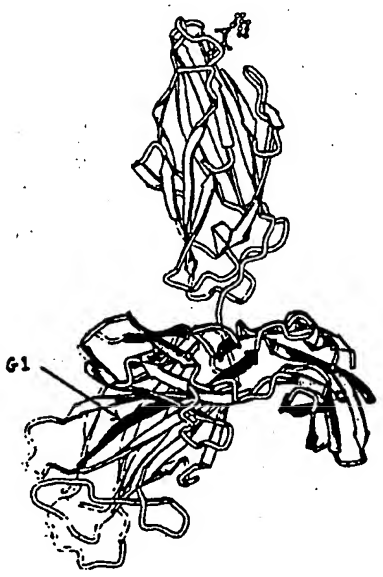
(72) Inventors: HULTGREN, Scott, J.; 1068 Polo Downs, Town & Country, MO 63017 (US). PINKNER, Jerome, S.; 658 North Forest, Webster Groves, MO 63119 (US). SAUER, Frederic; 7510 Parkdale, St. Louis, MO 63105 (US). BARNHART, Michelle; 4933 Potomac, St. Louis, MO 63139 (US). WAKSMAN, Gabriel; 500 West Drive, University City, MO 63130 (US). KNIGHT, Stefan; Noreens Vag 51, S-752 63 Uppsala (SE).

(74) Agents: GRANT, Alan, J. et al.; Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart & Olstein, 6 Becker Farm Road, Roseland, NJ 07068 (US).

(81) Designated States *(national)*: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

[Continued on next page]

(54) Title: DONOR STRAND COMPLEMENTED PILIN AND ADHESIN BROAD-BASED VACCINES



(57) Abstract: Modified polypeptides containing a pilus-protein sequence and a donor complementary sequence are disclosed, as well as complexes formed of a modified polypeptide and a pilin, or pilus-protein. Also disclosed are methods of using these novel polypeptides as a means of preventing and/or treating bacterial induced diseases, especially those caused by enterobacteria such as *Escherichia coli*. Methods of employing

[Continued on next page]

WO 01/04148 A2

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

these modified pilus-derived polypeptides and complexes as vaccines and for generating antibodies for further study as well as for clinical purposes are also disclosed herein. In addition, processes for large scale production of antibacterial vaccines containing said polypeptides and complexes are also described.

D scription

DONOR STRAND COMPLEMENTED PILIN AND ADHESIN BROAD-BASED VACCINES

This application claims the priority of U.S. Provisional Application 60/184,442, filed 23 February 2000, and of U.S. Provisional Application 60/144,359, filed 16 July 1999, and of U.S. Provisional Application 60/143,582, filed 13 July 1999, the disclosures of all three of these applications being hereby incorporated by reference in their entirety.

FIELD OF THE INVENTION

The present invention relates to immunogenic polypeptides and complexes useful in preventing and treating bacterial diseases and to methods of preparing such polypeptides and complexes.

BACKGROUND OF THE INVENTION

Newly formed protein chains can quickly fold *in vitro* to form the native conformation, a process often requiring no outside assistance and wherein the steric information for the three dimensional structure of the protein resides in the amino acid sequence. ATP-dependent proteins, called "chaperones," may aid folding of some proteins *in* vivo.

Conversely, there is another class of chaperones not requiring ATP for their functioning. These include the Pap-D-like periplasmic chaperones found in bacteria. As used herein, and unless expressly stated otherwise, the term "chaperone" means exclusively this latter class of periplasmic chaperones

having the further distinguishing characteristics as described below.

In bacterial species, this latter class of chaperones is responsible for mediating the synthesis of large scale oligomeric structures, for example, the construction of pili, the adhesive fibers expressed in most bacteria of the *Enterobacteriaceae* family (e.g., *Escherichia coli*).

Pili are heteropolymeric structures that are composed of several different structural proteins required for pilus assembly. Pili, also called fimbriae or fibrillae, facilitate the adhesive qualities of bacteria that often lead to colonization and infection of various tissues of the host animal, especially on mucosal surfaces. Such adhesion is facilitated by the presence in the pilus of a protein called an "adhesin," of which FimH is an example.

Different types of pili have been recognized. Type 1 pili-carrying bacteria recognize and bind to D-mannose in glycolipids and glycoproteins of bladder epithelial cells. Proteins forming the pili have been considered good candidates for vaccines. P pili are adhesive organelles encoded by eleven genes in the *pap* (pilus associated with pyelonephritis) gene cluster found on the chromosome of uropathogenic strains of *E. coli*. The biogenesis of P pili and Type 1 pili occurs via the highly conserved chaperone/usher pathway. (Thanassi et al, *Curr. Op. Microbiol.* 1,223 (1998); Hung et al, *EMBO J.* 15, 3792 (1994)).

Type 1 pili are composite fibers consisting of a short thin tip fibrillum joined to a thicker, rigid pilus rod. (C.C. Brinton, Jr., *Trans N.Y. Acad. Sci.* 27, 1003 (1965); Jones et al, *Proc. Natl. Acad. Sci. USA* 92, 2081 (1995)) The pilus fiber is an ordered array of homologous pilins (FimA, FimF, FimH, and FimG) with the FimH adhesin at its tip. FimH and FimG have been purified as a complex and comprise the bulk of the tip fibrillum, which may also contain FimF. (Jones et al (1995)) The rod is comprised of FimA monomers arranged in a right-handed helical cylinder (Brinton (1965)) Genes important in type 1 pilus biogenesis (*fimA-fimH*) are organized into the *fim* operon (Orndorff and Falkow, *J. Bacteriol.* 159, 736 (1984)) More specifically, FimH mediates binding to mannose-oligosaccharides present on mucosal surfaces. Thus, FimH mediates

adherence to mannosylated receptors on th bladd r pithelium and is critical to the ability of uropathogenic *Escherichia coli* t cause cystitis. (See: Mulvery et al, *Science* **282**, 1494 (1998); Langermann et al, *Science* **276**, 607 (1997); Krogfelt et al, *Infect. Immun.* **58**, 1995 (1990); Connell et al, *Proc. Natl. Acad. Sci. USA* **93**, 9827 (1996).

The PapD-like superfamily of periplasmic chaperones directs the assembly of over 30 diverse adhesive surface organelles that mediate the attachment of many different pathogenic bacteria to host tissues, a critical early step in the development of disease. (See Soto and Hultgren, *J. Bacteriol.* **181**, 1059 (1999)) PapD, the prototypical chaperone, is necessary for the assembly of P pili (Lindberg et al, *J. Bacteriol.* **171**, 6052 (1989)) whereas its homologue, called FimC, directs the assembly of type 1 pili (Jones et al, *Proc. Natl. Acad. Sci. USA* **90**, 8397 (1993)).

E. coli is the most common pathogen of the urinary tract, accounting for greater than 85% of cases of asymptomatic bacteriuria, acute cystitis and acute pyelonephritis, as well as greater than 60% of recurrent cystitis, and at least 35% of recurrent pyelonephritis infections. Because of the high incidence, continued persistence, and significant expense associated with *E. coli* urinary tract infections, there is a need for a prophylactic vaccine to reduce susceptibility to this disease.

While many factors contribute to the acquisition and progression of *E. coli* urinary tract infections, it is widely accepted that colonization of the urinary epithelium is a required early step. Therefore, disruption or prevention of pilus-mediated attachment of *E. coli* to urinary epithelia may prevent or retard the development of urinary tract infections.

For example, type 1 pili, as noted, are believed to be important in initiating colonization of the bladder and inducing cystitis, whereas P pili are thought to play a role in ascending infections and the ensuing pyelonephritis. Thus, pili mediate microbial attachment to the surfaces of cells, often the essential first step in the development of a disease, by binding to receptors

present in host tissues.

However, a major disadvantage to pilus-based vaccines has been the fact that the major immunodominant components of pilus fibers are often highly variable antigenically and therefore afford protection against only a limited number of bacterial strains. In contrast, pilus associated adhesins, such as FimH, are relatively conserved proteins among different species and strains of bacteria. Thus, FimH is relatively conserved not only among uropathogenic strains of *E. coli*, but also among a wide range of gram-negative bacteria. For example, many members of the family Enterobacteriaceae produce FimH and vaccines incorporating the FimH antigen should exhibit a broad spectrum of protection.

The major drawback to adhesin based vaccines of any kind has been the fact that adhesins are often only a minor component of the pilus, cannot be produced in large quantities, and therefore will tend not to elicit a particularly strong immunogenic effect. Although recombinant technology has succeeded in producing adhesin proteins in pure form, these are often rapidly proteolytically degraded when the corresponding chaperone is absent. Such adhesins are readily stabilized by the presence of periplasmic chaperone molecules (the latter also being important in proper synthesis of adhesins).

Gram negative bacteria, of which *Escherichia coli* is an example, have a characteristic cell surface arrangement. They possess an inner plasma membrane, surrounded by a peptidoglycan cell wall, which in turn is surrounded by an outer membrane, the latter being highly permeable to many substances. Between the cell wall and the outer membrane lies the periplasmic space. Proteins destined for secretion or assembly across the outer membrane often must fold within the periplasmic space prior to their secretion and/or assembly. Chaperones are often to be found within this periplasmic space. Among the proteins found in the periplasm are the adhesin FimH and its chaperone FimC.

Throughout this disclosure the terms pilus, pili, fimbrium, fimbriae,

fibrillum and fibrilla are be used interchangeably, with incidental use f the singular or plural form of any of these terms in no way limiting the breadth f the disclosed invention.

A "periplasmic chaperone" is defined herein as a protein localized in the periplasm of bacteria that is capable of forming complexes with a variety of proteins, especially pilus-proteins, including adhesins, especially FimH (where the corresponding chaperone is FimC) via recognition of a common binding epitope (or epitopes). Such chaperones are characterized by their similarity in properties to PapD, especially by their possession of an immunoglobulin-like fold for binding to pilus-proteins, such as adhesins. Chaperones serve as templates upon which proteins exported from the bacterial cell into the periplasm fold into their native conformations, especially where such proteins are intended to form oligomeric structures such as pili. Association of the chaperone-binding protein with the chaperone also serves to protect the binding proteins from degradation by proteases localized within the periplasm, increases their solubility in aqueous solution, and leads to their sequentially correct incorporation into an assembling pilus.

X-ray diffraction studies have helped to reveal the different domains found within the structures of the various adhesin and chaperone proteins. Resolution of the crystal structure of PapD, the prototype periplasmic chaperone, has revealed at least two domains having the overall topology of an immunoglobulin fold. [Holmgren & Branden, *Nature* **342**, 248 (1989)] The two domains are connected by a hinge region and oriented such that a cleft is created between these two domains. Further, other work has suggested that invariant surface-exposed residues that protrude into this cleft make up the subunit binding pocket. [Slonim et al, (1992) *EMBO J.* **11**, 4747-4756] Unlike cytoplasmic chaperones, PapD maintains target structures in native-like conformations [Lecher et al, (1989) *EMBO J.* **8**, 2703-2709; Kuehn et al, (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10586-10590] Such periplasmic chaperones have an effector function, specifically targeting the subunits to outer membrane assembly sites for their incorporation into pili and are characterized in part by the presence of an immunoglobulin-like fold.

Two of the genes in the *pap* operon – *papD* and *papC* – encode the chaperone and usher, respectively. Six genes encode structural pilus subunits, PapA, PapH, PapK, PapE, PapF, and PapG, which assemble into a heteropolymeric surface fiber with an adhesive tip (PapG). (Hultgren et al, *Cell* 73,887 (1993). The ability of PapG to bind to galabiose receptors in the human kidney is a critical event in the pathogenesis of pyelonephritis. The pilus consists of two major sub-assemblies, a thick, rigid rod made up of repeating PapA subunits arranged in a right-handed helical cylinder and a thin, flexible tip fiber (the tip fibrillum) extending from the distal end of the rod and composed primarily of repeating PapE subunits arranged in an open-helical configuration. Two components of the tip fibrillum, PapK and PapF, act as adaptors. PapK is thought to link the pilus rod to the base of the tip fibrillum and regulates its length: its incorporation terminates its growth and nucleates the formation of the pilus rod. PapF is thought to join the PapG adhesin to the distal end of the flexible tip fibrillum.

Like PapD, FimC uses its immunoglobulin-like domains to recognize and bind to pilus subunit proteins, such as the adhesin FimH.

The co-ordinated assembly of pili, as well as of other complex hetero-oligomeric organelles, requires correct incorporation of individual subunits in a predefined order during biogenesis and the prevention of premature associations between the intrinsically aggregative subunits. Type 1 pilus biogenesis proceeds via a highly conserved pathway that is involved in the assembly of over 30 adhesive organelles assembled by the adhesin-usher pathway in gram-negative bacteria. [Soto & Hultgren, *J. Bacteriol.* 181, 1059 (1999)] The assembly machinery is comprised of two specialized classes of proteins, a periplasmic chaperone and an outer membrane usher. [Thanassi et al, *Curr. Opin. Microbiol.* 1, 223 (1998)] Using its immunoglobulin-like folds, the periplasmic chaperone FimC forms periplasmic complexes with each of the pilus subunits prior to their incorporation into the pilus. [Soto & Hultgren (1999)] Furthermore, genetic and structural studies have shown that chaperones recognize a highly conserved motif present in the C-terminal

portions of all subunits assembled by all PapD-like chaperones. [Hung et al., *EMBO J.* 15, 3792 (1996); Kuehn et al., *Science* 262, 1234 (1993); Hultgr n et al., in *Molecular Biology of Chaperones and Folding Catalysts: Regulation, Cellular Functions and Mechanisms.* B. Bakau, Ed. (Harwood Academic Publishers, Amsterdam, 1999), p. 661]

The chaperone activity of FimC has been demonstrated and FimC has been shown to bind FimH, the adhesin of type 1 pili, to form periplasmic preassembly complexes. Thus, a FimH-FimC complex has been isolated using mannose-Sepharose chromatography of periplasm and then eluting with D-mannose. [see Jones et al (1993)] FimH is folded in the FimH-FimC complex in such a way that the mannose binding domain is in a native state and accessible for substrate binding. In addition, the amino acid sequence of FimC is known. [see Jones et al (1993)]

The crystal structure of PapD, the prototypical periplasmic chaperone, has been solved (Holmgren and Branden, *Nature* 342,248 (1989) and refined to 2.0 Å resolution, revealing a molecule with two immunoglobulin-like domains oriented in an L shape to form a cleft at their surface. All 30+ members of the periplasmic chaperone superfamily have a conserved hydrophobic core that maintains the overall features of the two domains. During pilus-biogenesis, PapD binds to and caps interactive surfaces on pilus subunits and prevents their premature aggregation in the periplasm. A combination of genetic, biochemical and crystallographic data has demonstrated that the G1 β-strand of PapD forms a β-zipper interaction with the highly conserved COOH-terminal motif of pilus subunits. The COOH-terminal motif also comprises at least part of a primary surface for subunit-subunit assembly interactions, indicating that the direct capping of a primary assembly surface is part of the molecular basis by which periplasmic chaperones prevent the premature oligomerization of pilus subunits. In addition, the β-zipper interaction has been proposed to facilitate the folding of the subunit into a native-like conformation via a template-mediated mechanism. (Soto et al, *EMBO J.* 17: 6155 (1998).

Whil the utility of adhesins as vaccines has been dem nstrat d, large scale production of adhesins and ther pilus-deriv d proteins has been complicated by the requirement of a chaperone that must be co-expressed with the adhesin in order for it to properly fold and result in a stable structure. It would therefore be highly advantageous if adhesins could be produced in pure form without the need of co-expressing the chaperone, and without the need for the chaperone, or any other protein, at all, thereby permitting large scale production of pure adhesins, or any other pilus subunits, for use, *inter alia*, as vaccines.

BRIEF SUMMARY OF THE INVENTION

The present invention relates to immunogenic complexes and polypeptides, comprising a pilus protein component or portion, and a donor strand component, or portion, wherein the pilus protein and donor strand may or may not be covalently bonded together.

It is an object of the present invention to provide polypeptides containing an amino acid sequence derived from a bacterial pilus-protein, including adhesins, such as FimH or a pilin such as PapK, and a portion of another protein acting as a donor complement, especially where the latter is a chaperone, such as FimC or PapD, or a portion of another subunit, such as FimG or FimF.

In specific embodiments, the present invention provides a polypeptide comprising a pilin protein, such as FimH, attached to a donor strand, such as a strand composed of the first 13 residues of FimG, separated by a short intervening sequence, such as a tetrapeptide, to form a single chain immunogenic polypeptide. Other specific embodiments include a similar structure employing PapK with a sequence of PapD attached to the C-terminus of PapK via a short amino acid linking sequence to form a single chain polypeptides. Of course, the invention also encompasses such structures wherein the pilin and donor strand segments are non-covalently linked to each

other s long as the correct conformation of the pilin is maintained.

In a highly preferred embodiment, polypeptides, or complexes thereof, of the present invention comprise a pilus subunit and a donor strand derived from the N-terminal extension of another pilus subunit.

It is still a further object of the present invention to provide vaccines containing the modified pilus polypeptides as a means of combating diseases, especially preventing diseases by vaccination, wherein such diseases are caused by bacterial species from which the adhesin structure is derived, especially *E. coli*.

It is an additional object of the present invention to provide for antibodies specific for the modified pilus-protein structures disclosed herein for use in treating diseases caused by bacteria from which the adhesin or pilin structure was derived, especially *E. coli*.

It is yet a further object of the present invention to provide a method for synthesizing modified pilus-proteins, in large scale, either in a free state or in a complex with another pilin protein, for use as vaccines, without the need of producing corresponding proteins, such as chaperones. Such methodology will thereby facilitate the production of such pilus-based (pilin-based and adhesin-based) vaccines as a single polypeptide chain, thus reducing time, cost and complexity of production and facilitate large scale vaccine production.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a comparison of the amino acid sequences of the various Fim proteins that compose the pili of bacteria such as *E. coli*. The sequences of FimH, FimA, FimF and FimG are all aligned to show corresponding sequences and domains. The end of the mannose-binding lectin domain and the start of the pilin domain in FimH are indicated by vertical

arrowheads above the sequences. The typ 1 pilin subunits (FimA, FimF and FimG) are aligned with the pilin domain of FimH using clustal W [see: Thompson et al, *Nucleic Acids Res.* 22, 4673 (1994)] and manually adjusted to minimize gaps in secondary structure elements. Gaps in the alignment are indicated by dots. Sequence numbering for FimH starts at position 22 in the pre-protein. Pilus subunits (including FimH) are expressed in the cytoplasm as pre-proteins with an amino terminal signal sequence that is cleaved during transport across the inner membrane. The first residue in FimH that is visible in our maps corresponds to Phe^{22} in the gene-derived sequence, which is the expected start of the FimH chain. To distinguish residues in the adhesin protein from residues in the chaperone, FimH residues will be denoted with an "H" and FimC residues with a "C" after the residue number. Residues involved in chaperone binding are indicated by an open circle above the residue. Residues in the carbohydrate binding pocket are boxed, with a large box marking the NH_2-terminal extensions in the pilin subunits. The conserved β-zipper motif found in all pilin subunits corresponds to the F β-strand. Limits and nomenclature for secondary structure elements are shown below the sequence.

Figure 2 shows the sequence of FimC. Residues involved in subunit binding are indicated by an open circle above the residue. Residues that are identical or conserved in all periplasmic chaperones are set against a darker background. Limits and nomenclature for secondary structure elements are shown below the residues.

Figure 3 shows β-sheet topology diagrams of the mannose binding domain (Fig. 3A) and the pilin domain (Fig. 3B) of FimH. The F strand is at the C-terminal end of the pilin domain and thus would appear at the C-terminus of the FimH molecule. Each strand depicted is readily correlated with the corresponding sequences of Figure 1 wherein the strand designation appears below the arrow indicating the residues making up that strand.

Figure 4 shows a molscript (P.J. Kraulis, *J. Appl. Cryst.*, 24, 946 (1991) ribbon diagram of the FimC-FimH complex, with FimH vertically and

FimC riented horizontally, as depicted by th arr ws (with th G1 strand indicated). The upper right shows a ball-and-stick representation of the C_HEGA molecule bound to the lectin domain of FimH and indicates the position of the carbohydrate-binding site at the tip of the domain.

Figure 5 shows a sequence alignment of P-pilus subunits (PapA, PapK, PapE, and PapF). The secondary structural elements of PapK are indicated above the aligned sequences, with β-strands and helices (including both α and 3_{10}) indicated. Residue numbers of PapK are indicated above the PapK sequence. Residues involved in contact with domains 1 and 2 of PapD are boxed. Residues strictly conserved among pilins are shown as the vertical quartets of G, C and Y residues (using standard one letter code) under residue numbers 7, 14, 54, 144 and 156.

Figure 6 shows the sequence and secondary structure definition of PapD. Residue numbers are indicated above the sequence, while secondary structural elements are indicated below it. Residues that contact PapK are boxed.

Figure 7 shows the topology of PapK with the nomenclature for secondary structures as described in Jones, *Curr. Op. Struct. Biol.* 3, 846 (1993). The β-strands are indicated as arrows, while helices are shown as cylinders. The inserting β-strand of PapD is indicated in the figure as G1.

Figure 8 shows the structure of the PapD-PapK complex and definition of secondary structure notation. Here, the molecular surface of PapK was calculated and displayed using GRASP [Nichols et al, *Proteins: Struct. Funct. Genet.* 11, 281 (1991)] The structure of PapD is shown as a ribbon. The insertion of the G1 β-strand of the PapD into a groove on the surface of PapK is shown.

Figure 9 shows the general arrangement for one embodiment of the present invention wherein a FimH has been linked to residues 1-13 of FimG using a tetrapeptide (DNKQ – using standard one letter amino acid code) linker

to form a donor strand complemented FimH (dscFimH). Th construction of this emb dim nt is described in detail in Example 1.

Figure 10 shows the results of an experiment in which C3H/HeJ mice were immunized with pilus-protein plus CFA/IFA. Here, a 4 week boost was provided IM (intramuscularly) and then the animals challenged at 9 weeks intraurethally with a dose of 11.3×10^7 cfu (colony forming units) of *E. coli* strain NU 14.

Figure 11 shows the results of an experiment in which C3H/HeJ mice were immunized with pilus-proteins plus MF59 with 4 and 16 week boosts IM (intramuscularly) with the indicated immunogens with endpoint titers shown at the left.

Figure 12 shows the results of an experiment in which C3H/HeJ mice were immunized with the indicated pilus-protein plus MF 59. Here, a 4 and 16 week boost was provided IM (intramuscularly) and then the animals challenged at 19 weeks intraurethally with a dose of 7×10^7 cfu (colony forming units) of *E. coli* strain NU 14.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is directed to polypeptides, especially immunogenic polypeptides, formed from a pilus-protein, such as, for example, an adhesin, and either a chaperone fragment or a pilin fragment, as well as dimeric complexes thereof, wherein the components of said complexes may or may not be covalently bound to each other. Such polypeptides, and complexes thereof, have the advantage of being synthesized in pure form and in a fully functional state without the need of accessory proteins, although the latter may be used to form specific complexes within the invention.

As used herein, the term "pilus-protein" means any protein or

polypeptide present in a pilus, specially a typ 1 pilus or a P pilus. This term nc mpasses all subunits f a pilus but xcludes haperones because, while they are integral in formation of a pilus, they are not incorporated into this organelle and thus are not subunits of it. Also as used herein, the term "adhesin" means a protein, especially a subunit of a pilus, with specific receptor binding properties, for example, FimH that has a lectin-binding domain for specifically binding mannose-bearing sites on cell surfaces, especially the surfaces of mucosal cells, most especially bladder cell of a mammal.

During pilus biogenesis, the chaperone, either FimC or PapD, binds to and forms stable complexes with individual pilus subunits. The chaperones consist of two immunoglobulin-like (Ig-like) domains oriented toward each other to form L-shaped molecules. [Holmgren and Brandan, *Nature* 342, 248 (1989); Pellucchia et al, *Nature Struct. Biol.* 5, 885 (1998)] The FimH adhesin has both a pilin domain and a receptor-binding domain. The PapK pilin and the pilin domain of FimH have Ig-like folds but lack the seventh C-terminal β-strand (strand G of Figure 3) present in canonical Ig-folds. The absence of this strand produces a deep groove along the surface of the pilin domain and exposes the hydrophobic core, thereby accounting for the instability of pilins when expressed without the chaperone (see examples hereinbelow). Thus, in the chaperone-subunit complex, the G1 strand of the chaperone completes an atypical Ig fold of the subunit by occupying the groove and runnig parallel to the subunit C-terminal F strand. In accordance with the present invention, this "donor strand complementation" interaction simultaneously stabilizes pilus subunits and caps their interactive surfaces, preventing their premature oligomerization in the periplasm. Also in accordance with the present invention, during pilus biogenesis, the highly conserved N-terminal extension of one subunit displaces the chaperone G1 strand from a neighboring subunit by a process referred to herein as "donor strand exchange." The N-terminal strand then inserts anti-parallel to the F-strand of the neighboring subunit to afford a mature pilus comprising an array of perfectly canonical Ig domains, each of which contributes a strand to the fold of its neighboring subunit.

Thus, further in accordance with the invention disclosed herein, the

contribution of a chap rone, such as FimC or PapD, to the overall structur of a pilin, such as in the FimC-FimH compl x, or in the PapD-PapK compl x, was determined by solving the structure of such complexes by X-ray diffraction [see: Choudhury et al, X-ray Structure of the FimC-FimH Chaperone-Adhesin Complex from Uropathogenic *E. coli*, *Science* **285**, 1061 (1999); Sauer et al, Structural Basis of Chaperone Function and Pilus Biogenesis, *Science* **285**, 1058 (1999); Barnhart et al., PapD-like Chaperones Provide the Missing Information for Folding of Pilin Proteins, *Proc. Natl. Acad. Sci. USA*, **10**, 1073/pnas.130183897 (published online June 20, 2000), the disclosures of all of which references are hereby incorporated by reference in their entirety]. For the FimC-FimH complex, the crystal structure was solved using MAD data to 2.7 Å collected from selenomethionyl FimC-FimH crystals. The crystals used for the structure determination belong to the symmetry group C2 with cell dimensions a = 139.08 Å, b = 139.08 Å, c = 214.49 Å, and β = 89.97 Å. The crystal structure revealed the presence of a pocket in the otherwise flat surface of the lectin domain. This pocket is large enough to accommodate a single mannose unit and is located at the tip of the domain, distal to the connection with the pilin domain. For FimH, the bottom of the pocket is defined by the N-terminus of the FimH molecule and is lined with typical carbohydrate binding side chains from Asn, Gln, and Asp residues in 3 loop regions.

This analysis showed that the FimH was folded into 2 domains of the all-beta class. These domains are aligned end to end so that the FimH molecule spans a length of over 100 Å. The amino terminal domain of the adhesin, FimH, contains the mannose binding domain, used to bind to the surface of a mucosal cell, and the C-terminal end forms the pilin domain, which is used to anchor the adhesin to the pilus. The NH_2-terminal mannose binding domain comprises residues 1H-156H while the C-terminal pilin domain, which is used to anchor the adhesin to the pilus, comprises residues 160H-279H (see Figure 1). In addition, a short extended linker (residues 157H-159H) connects the two domains.

As already stated, FimC has 2 immunoglobulin-like domains oriented at

about 90° to each other and with a deep cleft between the two domains. The pilin domain of FimH binds in the cleft of the chaperone but, with the seventh fold missing (see Figure 3), the hydrophobic core is exposed and the terminal portions of the pilin domain lie next to each other instead of at opposite ends of the pilin domain.

Thus, the pilin domain of FimH has the same topology as the NH_2-terminal domain of a number of chaperones but with the critical difference that the seventh strand of the fold is missing. A similar situation would occur in other pilus-proteins, such as FimG and FimF. In the FimH-FimC complex, the G1 strand (i.e., the seventh strand) of the chaperone is used to complement the pilin domain by being inserted between the second half of the A strand and the F strand of the pilin. Thus, the C-terminal, or F, strand of FimH (Figs. 1 and 3) forms a parallel beta pleated strand interaction with the G1 β-strand of FimC (Fig. 2) and has its COOH-terminal carboxyl group anchored in the crevice of the chaperone cleft through hydrogen bonds with the conserved residues Arg^{8C} and Lys^{112C} in FimC (superscripted numbers and letters refer to the residue number and to the identity of the protein involved – for example, "C" for FimC and "H" for FimH sequences). The residues involved in the FimH-FimC interactions are indicated in Figures 1 and 2 where the amino acid sequences of these proteins are shown.

The mechanism by which this process occurs, wherein the chaperone binds to the adhesin, here FimH, to direct proper folding of the adhesin is referred to herein as "donor strand complementation" and the strand that complements the fold or structure of the adhesin , such as the G1 β-strand of the chaperone FimC, is referred to as the "donor complement" segment or the "donor strand." More specifically, the G1 β-zipper strand of periplasmic chaperones contains a conserved motif of solvent exposed hydrophobic residues at positions 103, 105, and 107 in FimC (see Figure 2). In the FimC-FimH complex, these residues are thereby used to complete the unfinished hydrophobic core of FimH. In short, the pilin domain is an incomplete protein domain to which the chaperone donates its G1 strand so as to complete it. As a result, pilus-proteins, or pilins, as a class, are missing necessary steric

information needed for stability as well as f r f lding into the native 3 dimensional conformation and so require th pr senc f the chaperone, FimC in the case of FimH, in order to provide the necessary complementary, and compensating, structure.

The G1 strand of the chaperone simply inserts itself into the groove of FimH (the F and A2 line groove) so that engineering the same sequence at the C-terminal end of the pilin domain of FimH results in a FimH having the same structure (i.e., same overall shape) as if the chaperone were present. Alternatively, the N-terminal sequence of FimG may be employed since it is better for forming a completely canonical fold. In the same way, a donor strand can be engineered onto PapK by modeling the appropriate strand from the chaperone PapD or a strand from another suiteble pilus subunit structure.

Because this donor complemented structure of FimH, or any other adhesin, is necessary to the overall shape of the molecule and thus to the ability to stimulate the immune response, it is virtually impossible, in the absence of the chaperone, to generate such adhesins in a pure form to be used as immunogens, either by recombinant means or otherwise, and, thus, as a vaccine for prevention and treatment of infections.

In accordance with the present invention, the sequences missing from the structure of the putative-pilus protein immunogen are supplemented with appropriate donor strand sequences to take the place of the missing chaperone sequence. As disclosed herein, examples of such sequences and structures are provided. Also provided are rules providing guidance in facilitating the determination of appropriate sequences to use for donor complementing a wide variety of bacterial pilus proteins, especially where these are to be used in the development of compositions for preventing and/or treating bacterial diseases. Further, it is advantageous to supply these missing structures by recombinant means and thereby give rise to a fully active (i.e., immunogenic) adhesin molecule or complex.

FimA, FimF, and FimG all have a highly conserved extension of about

10-20 amino acids at the N-terminus compared t the FimH pilin domain. In th PapD-PapK structur , the PapK amino terminal extension is disordered and thus the first β-strand begins after the first cysteine, just as in the FimH pilin domain. In accordance with one embodiment of the present invention, an amino terminal extension of one subunit is used as a donor strand to provide the missing seventh strand in the neighboring subunit. Use of such neighboring subunits produces a complete canonical fold whereas the chaperone itself would complete an atypical fold.

Also in accordance with the present invention, crystallographic procedures likewise show that PapK has the same overall variable-region Ig-like fold as the amino terminal domain of PapD, with 2 β-sheets coming together in a β-sandwich. However, the Ig fold of PapK is incomplete in that it lacks the COOH-terminal seventh strand, G, which in canonical Ig folds forms an antiparallel B-sheet interaction with strand F and contributes to the hydrophobic core of the protein. It has now been found that in the PapD-PapK complex, this missing strand is provided by PapD, which donates its G1 β-strand to complete the Ig fold of PapK, in analogous fashion as was found for the FimH-FimC complex just described. However, the Ig fold produced thereby is atypical, since the donated strand runs parallel, rather than antiparallel, to strand F in PapK. The insertion of the G1 β-strand into the fold of the pilin, i.e., donor strand complementation, is similar to that observed in the crystal structure of the FimH-FimC complex.

The first 8 amino acid residues of PapK are disordered and the Ig fold of PapK begins with a short β-strand, A1 (see Figure 7), that makes typical anti-parallel hydrogen bonds with the COOH-terminal residues of strand B. This short β-sheet arrangement is interrupted by the insertion of the 3_{10} helical turn that results in strand A switching sides in the β-sandwich in order to make antiparallel B-strand interactions with the G1 strand of the chaperone. Strands A and B are connected by a short α-helix and it is strand B that forms the edge of one of the two β-sheets in the β-sandwich running antiparallel to strand E. Following strand B, the structure crosses over to the other side of the β-sandwich through a short 310 helix to form strand C1, which then runs

antiparallel to strand F.

The total buried surface area in the PapD-PapK complex is 3434Å^2 and there are two distinct sites on PapK that interact with two corresponding sites on PapD. Site K1 of PapK (see Figure 7) interacts with domain 1 of PapD (site D1 of Figure 6) and K1 contains a deep groove that runs the length of the subunit. The edges of this groove consist of strands A and F and its base is formed by the hydrophobic core of PapK. This groove is the result of the missing G β-strand in the Ig fold of PapK. Residues 101 to 112 (in site D1) of the G1 β-strand in the Ig fold of PapD (see Figure 6) insert into the K1 groove and make a β-zipper interaction with strand F of PapK (Figure 7) on one side of the groove. Residues 101 to 105 also make a β-zipper interaction with strand A2 on the other side of the groove. Insertion of the G1 β-strand also results in the formation of a continuous 5-stranded β-sheet which includes strands C1, F1, and G1 of PapD and the F and C1 strands of PapK. Alternating hydrophobic residues in the G1 β-strand of PapD interact with the hydrophobic base of the groove. Thus, donor strand complementation by the G1 β-strand of PapD shields the hydrophobic core of the pilin from exposure to the aqueous milieu of the periplasm. In addition, site K2 of PapK interacts with a site on the COOH-terminal domain (domain 2) of PapD. Thus, in the PapD-PapK complex structure, strand F of PapK forms one side of the groove into which the G1 β-strand of the chaperone inserts and is likely to assume the same structural role in pilins.

Genetic, biochemical and electron microscopic studies have shown that residues in two conserved motifs (the C-terminal F strand and an N-terminal motif) participate in subunit-subunit interactions necessary for pilus assembly. [See: Hung et al., *EMBO J.* 15, 3792 (1996); Kuehn et al., *Science* 262, 1234 (1993); Hultgren et al., in *Molecular Biology of Chaperones and Folding Catalysts: Regulation, Cellular Functions and Mechanisms.* B. Bakau, Ed. (Harwood Academic Publishers, Amsterdam, 1999), p. 661] An alignment of the pilin sequences (see Figure 1), based on the FimC-FimH crystal structure, revealed that the N-terminal motif was part of a 10-20 amino acid extension that was missing in the FimH pilin domain and disordered in the PapD-PapK

complex (a similar alignment for Pap-proteins is sh wn in Figure 4). This region contains a pattern f alternating hydroph bic residues similar to th G1 donor strand of the chaperone. As demonstrated by molecular modeling, the N-terminal extension of such a subunit takes the place of the G1 strand of the chaperone by fitting into the pilin groove. Thus, during pilus assembly, alternating hydrophobic side chains in the N-terminal extension can replace the hydrophobic side chains donated to the pilin core by the G1 strand of the chaperone, via the donor strand exchange mechanism of the present invention. The net result is that every subunit completes the immunoglobulin-like fold of its neighboring subunit.

The NH_2-terminal portion of pilins, corresponding to the disordered NH_2-terminus of PapK, forms an assembly surface on the pilin. The 8 NH_2-terminal residues are disordered in the Pap-D-PapK complex and protrude away from the main body of the structure, where they are free to interact with the groove of the preceding subunit located at the usher. In accordance with the present invention, therefor, the NH_2-terminus of an incoming subunit inserts into the groove of the preceding subunit, displacing the G1 β-strand of the chaperone (which process is facilitated by the usher). Such "donor strand exchange" implies that in the pilus, the NH_2-terminal strand of one subunit completes the immunoglobulin-like fold and thereby protects the hydrophobic core of the preceding subunit, much as the chaperone does in the periplasm.

The present invention is thus directed to pilus-proteins, including adhesins, such as FimH, or non-adhesins such as FimG and FimA, in which the missing fragment normally supplied by a chaperone, such as FimC, or a pilin via the donor strand exchange mechanism during pilus assembly, has been added to its amino acid sequence derived from that chaperone or pilin. In accordance with the invention, this result is accomplished by engineering a G1 β-strand of FimC (SEQ ID NO: 3 and 4), or an N-terminal extension strand of FimG (SEQ ID NO: 5 and 6) or FimF (SEQ ID NO: 7 and 8), or other similar sequence of related pilins, and in either a forward or reverse (i.e., inverted) sequence orientation onto the COOH-terminus of a pilus-protein, such as an adhesin, thereby removing the requirement of a chaperone and/or donor strand

complementation. For example, where the pilus-protein is an adhesin, such as FimH, the addition of th C-terminal β-strand to the adhesin complet s the immunoglobulin fold and thus removes the need for any type of accessory protein, such as a chaperone. Thus, by donating a secondary structural element to the fold of the pilin, for example, in the PapD-PapK complex, the chaperone not only contributes to the stability of the pilin but also prevents other pilins in the periplasm from binding to the groove of the chaperone-bound subunit. It should be noted that for P pili, PapG is the adhesin.

Due to topological requirements and constraints, and depending on the protein being engineered and the β-strand being "donated," the donor sequence can be engineered in a forward or reverse (i.e., inverted) amino acid sequence orientation with respect to the pilus-protein sequence being "complemented." However, because this donor strand must fold around and loop back onto the pilus strand in an anti-parallel orientation, such donor complemented structures can be engineered with the donor strand in a forward or reverse orientation. Whether enginered in a forward or reverse direction depends on the source of the sequence used to prepare the strand, for example, whether the donor strand is a match from the G1 strand of the corresponding chaperone or is modeled on the N-terminal strand of a neighboring subunit, such as where the N-terminal sequence of FimG is used to complement FimH. If engineered in a forward orientation, the N-terminal residue of the donor strand is bonded to the C-terminal residue of the pilus protein (e.g., FimH), while in the reverse orientation, the C-terminal residue of the donated strand is bonded to the C-terminal residue of the adhesin. Depending on the type of structural arrangement achieved, it may be necessary to also insert a linker sequence (perhaps as short as 4 residues or a structure of similar length and conformation) between the pilus-protein sequence and the donor strand sequence in order to afford sufficient flexibility for appropriate conformational folding necessary to attain the proper 3 dimensional structure (and such was done in carrying out embodiments of the present invention). Such engineering is readily accomplished either by direct chemical synthesis of the desired adhesin or by using a polynucleotide sequence in which the order of the codons encoding the donated strand

sequence have the desired orientation. Such donated strand sequence may, of course, be derived from the G1 strand of FimC or PapD, or the N-terminal extension strand of a pilus subunit, such as FimG. It should be mentioned that because the chaperone is not incorporated into the assembled pilus, such chaperones are not "pilus-proteins."

The ability to produce a fully formed and functional pilin protein without the need to provide for the chaperone, either by addition or as a co-expressed protein, is highly advantageous in that it permits large scale production of a single polypeptide chain, thereby reducing the time, cost and complexity of such production.

In furtherance of this objective, the present invention is directed to modified pilus-polypeptides, i.e., polypeptides or proteins derived from pili, especially type 1 pili and P pili, and forming subunits thereof, including adhesins such as FimH, and pilins such as PapK, comprising a pilus-protein portion and a donor complement portion all within the same amino acid sequence. Thus, with respect to a polypeptide within the invention the distinction between a donor complement portion and an adhesin or pilin portion is merely for convenience and in fact only a single polypeptide is contemplated herein, although such donor complement fragment may certainly be chemically attached to the adhesin portion by some chemical linkage other than a conventional peptide bond. Such modified proteins include FimH and PapK following donor strand complementation. Thus, donor strand complemented FimH would be denoted dscFimH (where dsc stands for "donor strand complemented"). In an alternative embodiment of the present invention, the donor strand need not be covalently attached at all but can merely be bonded non-covalently as part of a complex the overall shape of which is the native shape of the adhesin or pilin protein, such as FimH.

The problem of producing pure pilins or adhesins, free of a chaperone complex, and available for use as a vaccine, is solved by the present invention by providing any desired pilus protein (a term that does not include chaperones) from any bacterial structure assembled by the chaperone-usher

pathway (s Hung and Hultgren, *J. Struct. Biol.* 124, 201 – 220 (1998) for a review of this pathway) whose fold is complemented by a peptide, or even a non-peptide moiety, to form one structure or complex that is readily prepared in pure form and without the need for accessory proteins, such as chaperones.

In accordance with the disclosure herein, the present invention relates to an immunogenic complex comprising a pilus protein component and a donor strand component, wherein said pilus protein component may or may not be covalently bound to the donor strand component. Thus, the pilus and donor components may be held together by non-covalent interactions, including, but not limited to, electrostatic interactions, hydrophobic interactions, including van der Waals forces and general entropic forces, London forces, and the like. Such pilus protein component may or may not be an adhesin. Where it is an adhesin, especially preferred adhesins are FimH and PapG. Other pilus proteins are those selected from the group consisting of FimF and FimA (both from type 1 pili) and PapK, PapA, PapF, PapE, and PapH (all from type P pili). Donor strand components comprise an amino acid, or other polymeric strand, that is able to substitute for the missing strand of the pilus protein. Such substitution is determined by resort to the rules disclosed herein for utilizing the disclosed strands from PapD, FimC and FimG to replace missing strands of PapK and FimH.

The present invention also relates to polypeptides, especially immunogenic polypeptides, comprising a pilus protein portion and a donor complement portion, both of which may be part of the same protein complex. The donor complement portion may be a bacterial pilin or adhesin, a portion thereof, or may be derived from a bacterial pilin or adhesin. The pilus-protein portion of such a novel polypeptide will commonly be formed from the amino acid sequence of a native pilus-protein molecule, such as that making up the normal pilus from a bacterial cell (including all of the pilus proteins, including adhesins, just mentioned for the immunogenic complexes of the invention). Such bacterial cells may be any bacterial cells that have pilus structures, preferably from the *enterobacteriaceae* family, most especially *Escherichia coli*.

In forming such polypeptides, especially immunogenic polypeptides, the missing strand of the pilus protein can be compensated for using any appropriate donor strands, regardless of the source, so long as said donor strands will provide a functional donor complement (wherein the term "functional" means that the donor complementary strand, when attached to the C-terminus of the adhesin sequence, will permit the adhesin to assume its native three dimensional conformation without the need for the entire chaperone, or other separate conformation-directing molecule, to be present). In general, any sequence may be used if it can be modeled to fit into the groove.

For example, where an adhesin sequence provides the pilus-protein portion of the polypeptides of the present invention said adhesin will most preferably be the FimH protein found in bacterial cells such as *E. coli*. Thus, the amino acid sequence of FimH might conceivably vary slightly from one strain to another so that all such sequences are contemplated by the polypeptides of the invention. Most preferably, the FimH sequence of SEQ ID NO: 1 forms the adhesin portion, segment or fragment of the polypeptides of the invention when the pilus-protein is an adhesin.

As used herein, the terms "portion," "segment," and "fragment," refer to a continuous sequence of residues, such as amino acid residues, which sequence forms a subset of a larger sequence. For example, if a polypeptide were subjected to treatment with any of the common endopeptidases, such as trypsin or chymotrypsin, the oligopeptides resulting from such treatment would represent portions, segments or fragments of the starting polypeptide.

Where the donor complement portion is derived from a chaperone, especially a periplasmic bacterial chaperone, said chaperone is preferably FimC or PapD, and most preferably where the chaperone or pilin is derived from a bacterial cell of the family *enterobacteriaceae,* especially *E. coli,* and most preferably where the donor complement is the G1 β-strand of FimC or PapD (see Figure 2 and Figure 6, respectively, for location of G1 sequence) or an amino terminal extension sequence from a pilin such as FimG, especially where

said N-terminal sequence comprises at least.

The donor complement, or donor strand, portion of the polypeptides of the invention, or immunogenic complexes of the invention, may contain any amino acid, or other polymeric, sequence that provides appropriate compensation for the missing strand of the pilis protein. The appropriate strand to use as the donor complement is readily determined using the teaching of the present disclosure. Thus, to find a useful donor strand to complement a selected pilus protein, it is necessary to model the selected pilus protein with the appropriate chaperone and thereby determine the location and extent of the complementary sequence to use in forming the appropriate donor strand. Of course, the amino acid sequence of the relevant strands must be known.

Amino acid sequences for the pilus proteins (both pilins and adhesins) and chaperones recited herein are provided in Figures 1, 2, 5, and 6. The corresponding SEQ ID NOs are tabulated in Table 1.

Table 1.

Protein	Type Protein	Figure No.	SEQ ID NO.
FimH	Adhesin	1	1
FimC	Chaperone	2	2
FimG	Pilin	1	9
FimA	Pilin	1	12
FimF	Pilin	1	13
PapD	Chaperone	6	14
PapK	Pilin	5	15
PapA	Pilin	5	16
PapE	Pilin	5	17
PapF	Pilin	5	18
PapG	Adhesin		19

It should be kept in mind that the sequences provided for these proteins in the Sequence Listing can vary from species to species and from strain to strain within a species so that the sequences provided herein represent only one particular polypeptide. For example, SEQ ID NO: 19, for PapG, is for PapGII, which specifically binds sugars on human epithelial cells. The latter sequence also shows the signal sequence for secretion (which comprises the first 20 residues at the N-terminal end).

It should also be clearly stated that the present invention is equally operable with any proteins, and not just similar proteins, and regardless of the organism, so long as a desired conformation of the protein, preferably but not necessarily the native one, is achieved by donor strand complementationas disclosed herein. Thus, the mechanism of the present invention is directed to proteins that require a strand donated by another protein to achieve a given conformation, such as the native conformation, and is not necessarily restricted to use with the bacterial proteins used herein to describe the present invention. Thus, the mechanism described herein may be a general one.

In selected embodiments of the present invention, an appropriately selected donor strand is engineered onto the pilus protein to be complemented, either by synthesizing said pilus protein with the donor strand as part of the amino acid sequence, or by expressing a gene encoding the complemented structure (the so-called "dsc-pilus protein"), or by forming some other attachment, with the only requirement being that such dsc-pilus protein have the native three dimensional shape characteristic of the pilus protein as it occurs within the pilus, a structure stabilized *in vivo* by the presence of the chaperone. Such is readily determined by modeling according to the procedure described in the examples provided herein and in the description of the relevant protein-chaperone complexes disclosed hereinabove.

In one embodiment of the present invention, a donor strand amino acid sequence engineered at the C-terminal end of FimH, or other pilus-protein, using the 15-mer (or 15-residue) G1 fragment of FimC, with this G1 segment attached to the C-terminus of FimH using a linker sequence of from 1 to 20

amino acids in length, preferably about 3 t 10 r sidu s in length, with about 4 r sidues being an especially preferred embodiment. In another specific embodiment, the amino terminal extension of a pilin, especially FimG, is used as the donor strand. When such an N-terminal sequence of a pilin is used to form the donor strand, such a sequence normally comprises a sequence of about 6 or more amino acids, preferably about 6 to 20 amino acids, most preferably about 8 to 18 amino acids, especially 8 to 17 amino acids, where said sequence is identical to, or derived from, any of the N-terminal extensions of pilus subunits assembled by PapD-like chaperones. Such sequences could have amino acid substitutions at one or more positions to enhance stability, solubility, or other properties. Thus, the present invention encompasses any such sequences, especially those with two or more alternating hydrophobic residues, that completes the fold of and thus stabilizes the pilus subunit. In short, any sequence that accepts the donor strand complement can be complemented by the strands of the invention herein and any strand that completes the groove in the recipient protein to achieve the natural conformation can be used. The usefulness of such sequences will be most advantageous where one or more of the amino acids of the native G1 sequence have been replaced by amino acids of the same character (hydrophobicity, acidity, basicity, etc.). Consequently, the donor strands useful within the present invention do not have to be derived from a natural source but may be wholly synthetic in nature and character so long as they complement the recipient protein to produce the desired overall conformation, regardless of the use to which the donor strand complemented protein is to be put (so that it need not be for use as a vaccine but may have some other utility).

In an especially preferred embodiment, the first 13 amino acids of FimG (see Figure 1 and SEQ ID NO: 9) are linked to the C-terminal end of FimH (SEQ ID NO: 1) with the two sequences separated by a tetrapeptide linker composed of the sequence DNKQ (Asp-Asn-Lys-Gln) with the three segments

pilus protein – linker -- donor strand

linked together by conventional peptide bonds. The production of such a sequence (using FimH as pilus protein to form dscFimH) is provided in Example 1 below.

In addition, the novel polypeptides of the invention also include polypeptides having sequence homology, possibly less than 20% sequence identity, with the sequences of pilus proteins such as FimH, FimA, FimG, FimF, PapG, PapK, PapA, PapE, and PapF with the appropriate donor strand attached thereto to form the corresponding dsc-proteins with an optional linker sequence (for example, the N-terminal extension of the appropriate subunit fused to the corresponding protein to be complemented). For example, PapE and PapK have less than 20% sequence identity). The methods of the present invention are successful with any proteins assembled by the chaperone-usher pathway (i.e, assembly of a type 1 pilus). In addition, the replacement of selected amino acids as a means of increasing antigenic ability, or to increase or decrease the extent of other properties, is considered well within the skill of those in the art. Because the presence of FimC is not required to prepare the polypeptides of the invention, selected segments of the FimH portion that may be required for interaction and/or binding with FimC and serve no other purpose may advantageously be either eliminated or replaced with other amino acids, even those of different character. It is therefore deemed well within the skill of those in the art to make appropriate substitutions that may increase solubility, or any other desirable property, without sacrificing antigenicity.

It should be noted that, in accordance with the present invention, the term "percent identity" or "percent identical," when referring to a sequence, means that a sequence is compared to a claimed or described sequence after alignment of the sequence to be compared (the "Compared Sequence") with the described or claimed sequence (the "Reference Sequence"). The Percent Identity is then determined according to the following formula:

$$\text{Percent Identity} = 100\,[1-(C/R)]$$

wherein C is the number of differences between the Reference Sequence and

the Compared Sequenc over th length f alignment between the Reference Sequence and the Compared Sequence wh rein (i) each base or amino acid in the Reference Sequence that does not have a corresponding aligned base or amino acid in the Compared Sequence and (ii) each gap in the Reference Sequence and (iii) each aligned base or amino acid in the Reference Sequence that is different from an aligned base or amino acid in the Compared Sequence, constitutes a difference; and R is the number of bases or amino acids in the Reference Sequence over the length of the alignment with the Compared Sequence with any gap created in the Reference Sequence also being counted as a base or amino acid.

If an alignment exists between the Compared Sequence and the Reference Sequence for which the percent identity as calculated above is about equal to or greater than a specified minimum Percent Identity then the Compared Sequence has the specified minimum percent identity to the Reference Sequence even though alignments may exist in which the hereinabove calculated Percent Identity is less than the specified Percent Identity.

The present invention is not limited to singly engineered proteins but includes also engineered protein complexes. A non-limiting example of such a complex would be a complex formed between FimH and donor complemented FimG. In such a complex, FimH (i.e., a native FimH without the donor strand added and thus having an exposed tail groove) would be complexed with a FimG. The amino terminal extension of the latter pilus-protein would then insert into the exposed tail groove of FimH but would be prepared by the methods disclosed herein having a G1 β-strand of FimC, or other suitable donor complement strand, such as the N-terminal extension of FimF, attached to the FimG C-terminus (with the donor strand in its forward or reverse sequence orientation). These proteins would thereby be combined to form a self-assembling FimG-H complex, such as a dimer, especially a heterodimer, useful in an immunogenic composition, such as a vaccine. Donor complemented FimA (dscFimA) may combine with FimH in the same way that FimG does and forms a similar complex of an adhesin and a donor-complemented pilus-protein.

An especially preferred embodiment of the present invention comprises a polypeptide having an amino acid sequence comprising the amino acid sequence of FimH (or some variant thereof) as the pilus-protein portion and the sequence of the donor complement, such as the G1 β-strand of FimC or the amino terminal extension of FimG, especially where the G1 sequence of FimC (but not the amino terminal extension sequence of FimG) is inverted and attached through an amino acid linker to the C-terminal of FimH. Such linker may be composed of different amino acids, especially sequences capable or readily forming a loop structure so as to cause the donor strand to loop back toward the pilus protein and form an anti-parallel structure in place of the missing strand. Such a preferred embodiment is presented as SEQ ID NO: 10 wherein the linker is alternating serine/glycines and the donor strand is from FimC (SEQ ID NO:3). A similar preferred embodiment with FimH as the pilus-protein and the N-terminal extension of FimG separated by a DNKQ linker sequence is shown in SEQ ID NO: 11 (prepared in Example 1 below). An embodiment with PapG (SEQ ID NO: 19), DNKQ linker, and donor strand from PapF (SEQ ID NO: 20) is shown as SEQ ID NO: 21.

Donor strand sequences according to the present invention include the following:

G1 β-strand of FimC:

Forward:	NH_2-TLQLAIISRIKLYYR-COOH	SEQ ID NO: 3
Reverse:	NH_2-RYYLKIRSIIALQLT-COOH	SEQ ID NO: 4

N-Terminal sequence of PapF:

NH_2-DVQINIRGNVYIP-COOH	SEQ ID NO: 20

G1 β-strand of PapD:

Forward:	NH_2-DVTITVNGKVVAKP-COOH	SEQ ID NO: 5
Reverse:	NH_2-PKAVVKGNVTITVD-COOH	SEQ ID NO: 6

N-terminal extension of FimG:

NH2-DVTITVNGKVVAK-COOH SEQ ID NO: 7

N-terminal extension of FimF:

NH2-DSTITIRGYVRDN-COOH SEQ ID NO: 8

In keeping with the spirit and flexibility of the disclosure herein, any sequence that stabilizes the pilin or adhesin can be utilized advantageously as the donor strand or donor complement. As shown in Example 2, below, the donor strand complemented FimH (dscFimH) produced according to this particular embodiment was expressed in the absence of the chaperone in the periplasm and exhibited the properties of native FimH.

Regardless of the particular donor strand used as the complement in forming the dsc-pilus-protein structure, several structural requirements must be clearly borne in mind and may be considered "rules" for the selection of a suitable donor strand for donor complementation of a selected pilus protein according to the invention herein. For one thing, the donor strand binds to the pilus-protein structure in an anti-parallel arrangement so that it will commonly be preferred to attach the donor strand from a chaperone at the C-terminus, for example, of an adhesin, in a reverse orientation so that when it bends back it permits an anti-parallel orientation or in a forward direction in the case of an N-terminal sequence of a pilus subunit as donor strand, again to facilitate the desired anti-parallel orientation. In addition, because any such donor strand, regardless of source, should be permitted to bend back in order to form the appropriate strand complementation activity, the donor strand and the C-terminus of the pilus protein portion of a polypeptide of the present invention will usually be separated by an appropriate loop, or other suitable amino acid sequence. Such a looped linking structure could comprise a sequence of up to 20 amino acids, especially 1 to 10 residues, for example, about 4 or 5

residues, for example, the sequence of residues that occurs in the F1-G1 loop of the FimC chaperone, or especially useful would be the B1-C1 loop of PapD. Because the amino acid chain of the novel polypeptides of the invention are self-assembling, the donor strand will find its way to the appropriate complementation strand with no further guidance required, so long as sufficient conformational flexibility is provided to the engineered adhesin as a whole.

It is contemplated that the polypeptides of the present invention may be in isolated or purified form.

"Isolated" in the context of the present invention with respect to polypeptides (or polynucleotides) means that the material is removed from its original environment (*e.g.*, the cells used to recombinantly produce the polypeptides disclosed herein). Such peptides could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment. The polypeptides and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

The recombinant and/or immunogenic polypeptides, disclosed in accordance with the present invention, may also be in "purified" form. The term "purified" does not require absolute purity; rather, it is intended as a relative definition, and can include preparations that are highly purified or preparations that are only partially purified, as those terms are understood by those of skill in the relevant art. For example, polypeptides from individual clones isolated from a cDNA library have been conventionally purified to electrophoretic homogeneity. Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated. Furthermore, claimed polypeptides having a purity of preferably 0.001%, or at least 0.01% or 0.1%, and even desirably 1% by weight or greater is expressly contemplated.

For purposes of recombinantly producing the polypeptids of the invention, the term "expression product" means that polypeptide or protein that is the natural translation product of the gene and any nucleic acid sequence coding equivalents resulting from genetic code degeneracy and thus coding for the same amino acid(s).

Thus, the polypeptides of the present invention may also be present in the form of a composition. Such composition, where used for pharmaceutical purposes, will commonly have the polypeptide of the present invention suspended in a pharmacologically acceptable diluent or excipient.

The present invention is also directed to polynucleotides capable of coding for the polypeptides of the invention, especially polynucleotides encoding the amino acid sequence of SEQ ID NO: 11. Such polynucleotides would therefore contain at least one coding region for the polypeptides of the present invention, which would thus be an expression product thereof.

As used herein, the term "coding region" refers to that portion of a gene which either naturally or normally codes for the expression product of that gene in its natural genomic environment, i.e., the region coding *in vivo* for the native expression product of the gene. The coding region can be from a normal, mutated or altered gene, or can even be from a DNA sequence, or gene, wholly synthesized in the laboratory using methods well known to those of skill in the art of DNA synthesis.

In accordance with the present invention, the term "nucleotide sequence" refers to a heteropolymer of deoxyribonucleotides. Generally, DNA segments encoding the proteins provided by this invention are assembled from cDNA fragments and short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic gene which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon.

The term "expression product" means that polypeptide or protein that

is the natural translation product of th gen and any nucl ic acid sequence coding equival nts resulting from genetic code degeneracy and thus coding for the same amino acid(s).

The term "primer" means a short nucleic acid sequence that is paired with one strand of DNA and provides a free 3'OH end at which a DNA polymerase starts synthesis of a deoxyribonucleotide chain.

The term "promoter" means a region of DNA involved in binding of RNA polymerase to initiate transcription.

As used herein, reference to a DNA sequence includes both single stranded and double stranded DNA. Thus, the specific sequence, unless the context indicates otherwise, refers to the single strand DNA of such sequence, the duplex of such sequence with its complement (double stranded DNA) and the complement of such sequence.

The present invention is also directed to antibodies specific for, and antisera generated in response to, polypeptides of the invention. Such antibodies may be either polyclonal or monoclonal and may be generated, where monoclonal, from a cell, especially a hybridoma cell, by standard methods in the art. In addition, the present invention also relates to cells, and cell lines, genetically engineered to produce such antibodies after being transfected, or otherwise transformed, so that their genomes contain, within the main chromosome or as part of a plasmid or other vector, a polynucleotide encoding the genes for an antibody specific for a polypeptide of the invention, especially where said engineered cell is a cell capable of forming and secreting a fully formed antibody, such technology being known in the art.

The present invention also relates to vectors, such as plasmids, comprising the polynucleotides of the invention, said polynucleotides encoding polypeptides disclosed herein, and wherein such vectors are useful for transforming cells and permitting said transformed cells to express the polypeptides of the invention.

The present invention also relates to c lls transformed by such vectors and thereby expressing, with or without subsequent secretion thereof, of the polypeptides of the invention.

The present invention is also directed to vaccines and vaccine compositions comprising the polypeptides disclosed herein. Such a vaccine would comprise a composition containing an immunogenically effective amount of a polypeptide of the invention. A preferred embodiment of the invention is a vaccine composition comprising the polypeptide whose sequences are shown as SEQ ID NO: 10 and 11.

It is an object of the present invention to utilize an immunogenic composition for a vaccine (or to produce antibodies for use as a diagnostic or as a passive vaccine) comprising a bacterial polypeptide of the invention. In one embodiment, proteins and fragments (naturally or recombinantly produced, as well as functional analogs) from bacteria that produce type 1 or type P pili are contemplated. Even more particularly, *E. coli* is contemplated as the source.

In another aspect of the invention, such an immunogenic composition may be utilized to produce antibodies to diagnose urinary tract infections, or to produce vaccines for prophylaxis and/or treatment of such infections as well as booster vaccines to maintain a high titer of antibodies against the immunogen(s) of the immunogenic composition.

While other antigens have been utilized to produce antibodies for diagnosis and for the prophylaxis and/or treatment of bacterial urinary tract infections, there is a need for improved or more efficient vaccines. Such vaccines should have an improved or enhanced effect in preventing bacterial infections mediated by pilus proteins.

There is a need for improved antigenic compositions comprising adhesins and pilins for stimulating high-titer specific antisera to provide

protection against infection by pathogenic bacteria and also f r use as diagnostic reagents.

In one aspect, the present invention is directed to an immunogenic composition comprising a purified dsc-pilus protein polypeptide or immunogenic complex thereof. A specific embodiment comprises a native adhesin, preferably FimH, and a donor complement, such as one derived from a periplasmic chaperone, preferably FimC, most preferably the G1 strand of FimC, or an amino terminal extension of a pilin, preferably FimG, most preferably no more than the first 17 N-terminal residues of FimG, especially the first 13 residues thereof, with the dsc-pilus-protein maintained in the complex in an immunogenic form capable of inducing an immune response when appropriately introduced into a human or other mammalian species. Thus, a preferred embodiment is one that includes the N-terminal extension of another subunit.

The dsc-polypeptides and complexes of the present invention are primarily intended for use as vaccines. Generally, vaccines are prepared as injectables, in the form of aqueous solutions or suspensions. Vaccines in an oil base are also well known such as for inhaling. Solid forms which are dissolved or suspended prior to use may also be formulated. Pharmaceutical carriers are generally added that are compatible with the active ingredients and acceptable for pharmaceutical use. Examples of such carriers include, but are not limited to, water, saline solutions, dextrose, or glycerol. Combinations of carriers may also be used.

Vaccine compositions may further incorporate additional substances to stabilize pH, or to function as adjuvants, wetting agents. or emulsifying agents, which can serve to improve the effectiveness of the vaccine.

Vaccines are generally formulated for parenteral administration and are injected either subcutaneously or intramuscularly. Such vaccines can also be formulated as suppositories or for oral administration, using methods known in the art.

The amount of vaccine sufficient to confer immunity to pathogenic bacteria is determined by methods well known to those skilled in the art. This quantity will be determined based upon the characteristics of the vaccine recipient and the level of immunity required. Typically, the amount of vaccine to be administered will be determined based upon the judgment of a skilled physician. Where vaccines are administered by subcutaneous or intramuscular injection, a range of 50 to 500 µg purified protein may be given.

In addition to use as vaccines, the polypeptides of the present invention, and immunogenic fragments thereof, can be used as immunogens to stimulate the production of antibodies for use in passive immunotherapy, for use as diagnostic reagents, and for use as reagents in other processes such as affinity chromatography.

The present invention also provides for a recombinant production or synthesis of the proteins and polypeptides of the invention without the need for any chaperone being present or the need to co-express any chaperone during production of the replacement adhesin for use as a vaccine (or as an immunogen to produce antibodies for diagnostic or therapeutic purposes).

Recombinant polypeptides of the invention are readily produced by methods of genetic engineering already well known in the art (e.g., Example 1, below) or by direct synthesis by well known, even automated, methods. Therefore, a compendium of procedures for preparing the polypeptides and complexes of the invention need not be recited herein.

In addition to producing a genetically engineered or synthetic sequence for the complete adhesin-donor complement polypeptide, it is also possible to attach the appropriate donor strand fragment at or near the COOH-end of the adhesin chain by some chemical linker other than a conventional oligopeptide using a standard peptide bond. Such chemically fused structures are contemplated by the present invention, the nature of such structures being limited only by the imagination of chemists seeking to produce functional

polypeptides of th invention. Such linking structures also include standard polymers forming an appropriate looping structur r may be by any of the non-covalent interactions listed above.

Thus, in accordance with the present invention, the ability to engineer a recombinant adhesin-donor complement protein, a dsc-pilus polypeptide, without the requirement of co-producing a chaperone, or adding an exogenous chaperone, or fragments thereof, to the expression medium, to enable the completion of the adhesin or pilin native structure and conformation readily permits large scale synthesis of immunogenic polypeptides for use as vaccines.

The polynucleotides encoding the polypeptides of the invention may also have the coding sequence fused in frame to a marker sequence which allows for purification of the polypeptides of the present invention. The marker sequence may be, for example, a hexa-histidine tag supplied by a pQE-9 vector to provide for purification of the mature polypeptides fused to the marker in the case of a bacterial host, or, for example, the marker sequence may be a hemagglutinin (HA) tag when a mammalian host, e.g. COS-7 cells, is used. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein (Wilson, I., et al., Cell, 37:767 (1984)).

When host cells are genetically engineered (transduced or transformed or transfected) with the vectors comprising a polynucleotide encoding a polypeptide of the present invention, the vector may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the polynucleotides which encode such polypeptides. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

Vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids

and phage DNA, viral DNA such as vaccinia, aden virus, fowl pox virus, and pseudorabies. Howev r, any other vector may be used as l ng as it is replicable and viable in the host.

The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. As representative examples of such promoters, there may be mentioned: LTR or SV40 promoter, the E. coli. lac or trp, the phage lambda P_L promoter and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in E. coli.

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the proteins.

As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as *E. coli, Streptomyces, Salmonella typhimurium*; fungal cells, such as yeast; insect cells such as *Drosophila S2*

and Spodoptera Sf9; animal cells such as CHO, COS or Bowes m lan ma; adenovirus s; plant cells, etc. The selection of an appropriat h st is deemed to be within the scope of those skilled in the art from the teachings herein.

More particularly, the present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example. Bacterial: pQE70, pQE60, pQE-9 (Qiagen, Inc.), pbs, pD10, phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia). However, any other plasmid or vector may be used as long as they are replicable and viable in the host.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P_R, P_L and TRP. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

Host cells containing the above-described constructs can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or

electroporation (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)).

The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the polypeptides of the invention can be synthetically produced by conventional peptide synthesizers.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), Wu et al, *Methods in Gene Biotechnology* (CRC Press, New York, NY, 1997), and *Recombinant Gene Expression Protocols*, in *Methods in Molecular Biology*, Vol. 62, (Tuan, ed., Humana Press, Totowa, NJ, 1997), the disclosures of which are hereby incorporated by reference.

Transcription of the DNA encoding the polypeptides of the present invention by higher eukaryotes is increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act on a promoter to increase its transcription. Examples including the SV40 enhancer on the late side of the replication origin bp 100 to 270, a cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase

(PGK), α-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli, Bacillus subtilis, Salmonella typhimurium* and various species within the genera *Pseudomonas, Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM1 (Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further

purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, a french press, mechanical disruption, or use of cell lysing agents, such methods are well know to those skilled in the art.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell, 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

The polypeptides can be recovered and/or purified from recombinant cell cultures by well-known protein recovery and purification methods. Such methodology may include ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. In this respect, chaperones may be used in such a refolding procedure. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

The polypeptides that are useful as immunogens in the present invention may be a naturally purified product, or a product of chemical synthetic procedures, or produced by recombinant techniques from a prokaryotic or eukaryotic host (for example, by bacterial, yeast, higher plant,

insect and mammalian cells in culture). Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. Particularly preferred immunogens are FimH-β-strand polypeptides or mannose-binding fragments thereof since FimH is highly conserved among many bacterial species. Therefore, antibodies against FimH (or its mannose-binding fragments) should bind to FimH of other bacterial species (in addition to *E. coli*) and vaccines against *E. coli* FimH (or FimH mannose-binding fragments) should give protection against other bacterial infections in addition to *E. coli* infections (for example, against other enterobacteriacea infections).

The polypeptides, their fragments or other derivatives, or analogs thereof, or cells expressing them can also be used as an immunogen to produce antibodies thereto. These antibodies can be, for example, polyclonal or monoclonal antibodies. The present invention also includes chimeric, single chain, and humanized antibodies, as well as Fab fragments, or the product of an Fab expression library. Various procedures known in the art may be used for the production of such antibodies and fragments.

Antibodies generated against the polypeptides corresponding to a sequence of the present invention can be obtained by direct injection of the polypeptides into an animal or by administering the polypeptides to an animal, preferably a nonhuman. The antibody so obtained will then bind the polypeptides itself. In this manner, even a sequence encoding only a fragment of the polypeptides can be used to generate antibodies binding the whole native polypeptides.

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein, 1975, Nature, 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole, et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96).

Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to immunogenic polypeptide products of this invention. Also, transgenic mice may be used to express humanized antibodies to immunogenic polypeptide products of this invention. In addition, cells can be transformed with gene sequences corresponding to antibody chains containing variable regions complementary to the polypeptides of the invention and thereby generate engineered antibodies to the polypeptides disclosed herein.

In carrying out the procedures of the present invention it is of course to be understood that reference to particular buffers, media, reagents, cells, culture conditions and the like are not intended to be limiting, but are to be read so as to include all related materials that one of ordinary skill in the art would recognize as being of interest or value in the particular context in which that discussion is presented. For example, it is often possible to substitute one buffer system or culture medium for another and still achieve similar, if not identical, results. Those of skill in the art will have sufficient knowledge of such systems and methodologies so as to be able, without undue experimentation, to make such substitutions as will optimally serve their purposes in using the methods and procedures disclosed herein.

The present invention will now be further described by way of the following non-limiting examples. In applying the disclosure of these examples, it should be kept clearly in mind that other and different embodiments of the present invention will no doubt suggest themselves to those of skill in the relevant art.

Example 1

Folding of FimH *in cis* using a donor strand

It was shown experimentally that a pilus subunit protein, otherwise unable to fold independently (or to fold inefficiently) due to lack of a C-

terminal G β-strand in the absence of a periplasmic chaperonet provide the correct steric information, was able to fold correctly when the missing strand was provided *in cis*. Here, the missing seventh β-strand was fused onto the 3'end of FimH. Here, the DNA sequence encoding the first 13 amino acids of FimG (see Figure 1), referred to herein as the donor strand, was provided to FimH *in cis*, by fusing it directly to the 3'-end of the FimH coding sequence, with the resulting DNA sequence encoding the donor strand complemented FimH protein (called dscFimH).In addition, a hairpin loop region in PapD consisting of Asp-Asn-Lys-Gln was inserted upstream of the donor strand to form a hinge region in the expressed protein and thereby allow the donor strand to fold back and form an anti-parallel arrangement with FimH (with the arrangement shown in Figure 9).

To construct the dsc-FimH, the following two oligonucleotides were annealed together and ligated into the Cla1 and BamH1 sites of pUC18-FimH to create pUC18-dscFimH (here, with DNKQ as linker - the latter are the 4 amino acids used for the loop or hinge region, as described above, using standard 1 letter code for amino acids) where the top (coding strand) is:

5'-CGATTATTGGCGTGACTTTTGTTTATCAAGATAACAAACAGGATGTCA
CCATCACGGTGAACGGTAAGGTCGTCGCCAAATAAG-3'

SEQ ID NO: 22

and for bottom strand:

5'-GATCCTTATTTGGCGACGACCTTACCGTTCACCGTGATGGTGACCATC
CTGTTTGTTATCTTGATAAACAAAAGTCACGCCAATAAT-3'

SEQ ID NO: 23

Here, pUC18-dscFimH was sequenced followed by subcloning into the EcoR1 and BamH1 sites of pTrc99A (Amann et al, *Gene* 69, 301 (1988). to create pTrc-dscFimH. The *fimH* was then subcloned from pUC18-FimH into pTrc99A using the EcoR1 and BamH1 sites to create pTrc-FimH.

FimH, FimH+FimC, and dscFimH wer expressed separately. Th plasmids encoding the FimH (pTrc-FimH), dscFimH (pTrc-dscFimH), and FimC (pJP4) were expressed in C600 (see Jones et al, *Proc. Natl. Acad. Sci. USA,* **90, 8397 (1993)**). Overnight cultures were diluted 1:100 into Luria broth and grown to an OD_{600} of 0.6 followed by induction with 0.5 mM IPTG for 1 hour. Periplasms were prepared by a known procedure (see Slonim et al., *EMBO J.* **11**, 4747 (1992).

The presence of FimC and FimH or dscFimH in periplasmic extracts was monitored by immunoblotting using anti-FimCH antibodies. In addition, mannose-sepharose chromatography (Jones et al, 1993, above) was used to determine the ability of FimH or dscFimH to bind its receptor. FimH was degraded when expressed alone but was stabilized by co-expression of the chaperone. In contrast to FimH, dscFimH was stable in the periplasm in the absence of FimC. FimH bound to mannose-sepharose beads when it was co-expressed with FimC and thus eluted as a FimCH complex. Since FimH is degraded in the absence of FimC, no full length FimH eluted from the mannose-sepharose beads. In contrast, dscFimH bound to, and specifically eluted from, the mannose beads when expressed alone. When FimC was co-expressed with dscFimH it did not form a complex with dscFimH and thus did not co-elute with dscFimH from the mannose-sepharose beads.

In contrast to FimH, dscFimH was not able to complement a *fimH*⁻ (FimH negative) type 1 gene cluster to restore a hemagglutination positive phenotype. Here, *fimH* and *dscFimH* were subcloned from pUC18-FimH and pUC18-dscFimH into pBad18-Kn (see Guzman et al, *J. Bacteriol.*, **177**, 4121 (1995). using the EcoR1 and Xba1 sites to create pBad-FimH and pBad-dscFimH which were transformed into ORN103/pETS10. ORN103 does not produce type 1 pili and pETS10 encodes a *fimH* type 1 gene cluster. The strains were diluted 1:100 into Luria broth and grown to an OD_{600} of 0.8 followed by induction with 0.1 mM IPTG and 0.02% arabinose for one hour. The cells were harvested and hemagglutination assays were performed by methods known in the art (see: Hultgren et al, *Mol. Microbiol.* **4**, 1311

(1990).

In accordance with the present invention, the added donor strand occupied the groove and completed the Ig fold of the FimH pilin domain, thus shielding the surface that would normally interact first with the chaperone and then with another subunit in the pilus. Thus, dscFimH did not bind FimC nor assemble into the pilus.

FimH and dscFimH were purified and their denaturation curves were obtained. FimH was isolated from a FimCH complex by incubating the complex in 3M urea and subjecting it to cation exchange chromatography which yielded 2 peaks, one of which contained pure FimH. The FimCH complex was purified from the periplasm of C600/pHJ9205/pHJ20 (see: Jones et al (1993) and Slonim et al (1992, above). The periplasmic extracts were dialyzed against 20 mM MES pH 5.4, injected onto a Source 15S column (Pharmacia), and FimCH eluted with 65 mM NaCl. The eluate was injected onto a Butyl4FF column in 0.55M $(NH_4)_2SO_4$/20 mM MES pH 5.4 and FimCH eluted at 0.3 M $(NH_4)_2SO_4$. The FimCH complex was brought to 3M urea to separate the two subunits. Pure FimH in 3M urea was collected from the flow through of a Source 15S column. FimH retained its native structure in 3M urea as determined by circular dichroism (CD) and its ability to bind mannose.

DscFimH was also purified from the periplasm of C600/pTrc-dscFimH (see Slonim et al (1992)).The periplasm was dialyzed against 20 mM Tris-Cl pH 8.8 and dscFimH was collected from the flow through of a Source 15Q column. This flow through was injected onto a Butyl4FF column in 0.9M $(NH_4)_2SO_4$/20 mM Tris-CL pH 8.8 and dscFimCH eluted at 0.4 M $(NH_4)_2SO_4$. The eluate was then loaded onto a Source 15S column in 20 mM MES pH 4.7 and dscFimH eluted at 55 mM NaCl.

FimH and dscFimH had similar denaturation curves with denaturation complete only at concentrations above 8.5 M urea (+4mM DTT) as determined by tyrosine fluorescence spectroscopy emission maxima (350

nm). For this measurement, 22.5 µg of FimH or dscFimH in 20 mM MES pH 6.5 + 4 mM DTT was incubated with th appr priate ur a concentration. Fluorescence was measured using an excitation wavelength of 290 nM with emission at 350 nm on an AlphaScan PTI fluorometer. FimH did not begin to denature until a concentration of 6.5 M urea was reached, with the midpoint of the denaturation curve occurring at approximately 7.5 M urea.

Due to the twist in the β-sheet formed by strands D, C and F, the G1 strand of FimC is unable to satisfy all potential backbone hydrogen bonding interactions with the F strand of FimH. Modeling of the finished structures was performed using SYBYL (Tripos Associates) and Insight II (Molecular Simulations Inc.) running on a Silicon Graphics workstation.

Example 2

In Vitro Folding Assay

An *in vitro* folding assay was used to demonstrate that the missing steric information in the amino acid sequence of pilus subunit proteins can be provided *in cis*. This assay was based on attempted refolding of urea-denatured FimH (obtained from a FimCH complex as already described) and dscFimH as determined by examination of the CD spectra after rapid dilution of the denatured proteins. Spectra were measured from 150 µg of protein in 20 mM MES pH 6.5 using a 0.02 cm cell in a JASCO J715 spectropolarimeter. Denatured proteins were diluted to 0.45 M urea.

Before denaturation, FimH (3 M urea) had a virtually identical β-sheet CD spectrum as compared to dscFimH. After denaturation in 9 M urea (+4 mM DTT), the CD spectra of FimH and dscFimH became characteristic of

non-native pr teins. Light scattering of the denatured dscFimH indicated th presence of large aggregates. However, rapid dilution f dscFimH led t the refolding of the protein into its native β-sheet structure. The refolded dscFimH bound mannose and was monodisperse (as shown by light scattering) indicating that it had refolded into its native structure. In contrast, attempts to refold FimH led to insoluble aggregates and therefore elicited no signal after filtering. FimC was unable to bind to denatured FimH after its rapid dilution. However, if FimC was present in the diluent, FimH formed a complex with FimC and folded into its native mannose-binding β-sheet structure. Thus, in these assays, dscFimH folded independently whereas FimH folded in the presence of, but not in the absence of, FimC. FimC was capable of binding to native FimH separated from the FimCH complex by 3 M urea, confirming that the chaperone can indeed bind to folded subunits. As further evidence, a mutation in Arg 8 of FimC (see Figure 2), a residue critical in chaperone complex formation, abolished the ability of the mutant protein to bind to native FimH or to facilitate re-folding of denatured FimH.

In sum, these results show that the chaperone is necessary for subunit folding by providing the subunit protein with missing steric information. Here, the information required for subunit folding resides in two polypeptides. In such an arrangement, the C-terminal carboxyl group of the F strand of the subunit anchors to the conserved Arg8 and Lys112 residues in the chaperone cleft. Subsequent β-zippering along the G1 strand facilitates the formation of the initial F β-strand in the pilin, which in turn initiates β-sheet formation. These interactions would position the strand-F hydrophobic side chains of the subunit in register with the G1 strand alternating hydrophobic residues of the chaperone so as to facilitate the proper collapse of the hydrophobic core of the subunit. In dscFimH, the steric information normally provided by the chaperone is now present in a single polypeptide chain, provided by the sequence corresponding to the N-terminus of FimG. The missing sequence provided *in cis* enables the pilin domain of FimH to fold into a perfect canonical Ig fold, mimicking the fold that is otherwise completed by FimG in the tip fibrillum.

As already stated, FimH based vaccines hav been shown to protect mice and monkeys from xperimental bladder infections (see: Lang rmann et al, *Science* **276**, 607 (1997); Langermann et al, *J. Infect. Dis.* in press). However, the production of such vaccines required co-expression of FimC and the purification of a FimCH complex. In accordance with the invention disclosed herein, donor strand complementation now permits production of dscFimH vaccines. These vaccines are advantageous in that they comprise a single subunit and are expected to be more stable due to the anti-parallel donor strand. Thus, vaccination with dscFimH has been found to produce anti-FimH titers that are comparable to those achieved with the FimCH complex. Thus, the methods taught by the present invention readily facilitate the production of adhesin-based vaccines generally.

Example 3

dsc-Immunogen Dosages

This experiment evaluated dosages of the dsc immunogen (dose down) at 15 µg, 3 µg, and 0.6 µg. CFA/IFA was used as an adjuvant. Here, 30, 6 and 1.2 µg FimCH was used for comparison. It was expected that the ratio between the dsc and FimCH would be similar since the amount of FimH at each dose of FimCH complex (1:1 ratio of FimH to FimC) is equivalent to the amount present in each dose of dscFimH. Endpoint titers using anti-FimH T3 and Anti-FimdscFimH as detecting antigens were used to illustrate this correlation, evaluate immune responses and demonstrate comparability. With the exception of the 0.6 µg dose of the dsc immunogen (vs. 1.2 µg FimCH), end point titers appeared similar. Results are tabulated in Table 2 as endpoint dilutions using ELISA. Each group was done in duplicate. Note for example that Group 1 and Group 7 differ in the amount of pilin used as immunogen. This is because Group 7 received 30 µg FimCH, which is the complex of FimC and FimH, whereas Group 1 received 15 µg dscFimH, which is FimH with only the donor strand segment at the end. Thus, by doubling the amount of the FimCH complex a more closely related molar

quantity was achieved since the molecular weight of th complex is alm st twice that of the corresponding donor strand complem nt d material. This demonstrates the immunogenicity of the dsc protein versus the complex pilin-chaperone complex.

Table 2

		Anti-FimH T3		Anti-FimH dsc	
Group	Immunogen	3 week	8 week	3 week	8 week
1	15 µg FimH dsc	204800	1638400	204800	1638400
2	15 µg FimH dsc	102400	409600	204800	1638400
3	3 µg FimH dsc	102400	819200	102400	1638400
4	3 µg FimH dsc	51200	409600	51200	819200
5	0.6 µg FimH dsc	3200	25600	12800	409600
6	0.6 µg FimH dsc	400	25600	6400	409600
7	30 µg FimCH	204800	1638400	204800	1638400
8	30 µg FimCH	204800	1638400	204800	1638400
9	CFA/IFA	100	100	100	100
10	CFA/IFA	100	100	100	100
11	No Injection	100	100	100	100
12	No Injection	100	100	100	100
13	6 µg FimCH	102400	1638400	102400	1638400
14	6 µg FimCH	102400	819200	102400	1638400
15	1.2 µg FimCH	51200	819200	51200	1638400
16	1.2 µg FimCH	51200	819200	51200	1638400

The results of challenge with the NU 14 strain of *E. coli*, administered intraurethrally, is shown in Figure 10. The results indicate protection by the donor strand complemented protein (dscFimH). Here, the dsc antigen was compared to control (CFA/IFA - complete Freund's adjuvant/incomplete Freund's adjuvant) and naive (no adjuvant or antigen) as well as to high dose FimCH (CFA/IFA). Significant protection was observed with the dscFimH when compared to either the control or naive groups, at least at the 3 µg and 15 µg doses.

Example 4

Immunogenicity of DSC Protein

In these experiments, MF 59 (Chiron Corp.) was used as adjuvant and monitoring was continued over an 18 week period. Overall, the responses were comparable between the FimCH and dscFimH at all doses. Endpoint titers are shown in Table 3 as well as in the graphs of Figure 11. In the challenge experiment (results depicted in Figure 12), better protection was observed at the lower versus the higher doses (which could be due to the use of the MF 59 adjuvant). The 0.6 µg dose appeared to be the best.

Table 3.

Group	Immun gen	A	B	C	D	E
		3 week	8 week	12 week	15 week	18 week
1	15 µG FimH dsc	12800	204800	204800	51200	409600
2	15 µG FimH dsc	25600	409600	409600	102400	409600
3	3 µG FimH dsc	800	204800	204800	102400	409600
4	3 µG FimH dsc	6400	204800	204800	102400	409600
5	0.6 µgFimH dsc	100	12800	12800	6400	204800
6	0.6µG FimH dsc	100	51200	51200	12800	409600
7	30 µg FimCH	25600	409600	409600	204800	409600
8	30 µg FimCH	25600	409600	204800	102400	409600
9	MF 59	100	100	100	100	100
10	MF59	100	100	100	100	100
11	Naive	100	100	100	100	100
12	Naive	100	100	100	100	200
13	6 µg FimCH	25600	409600	409600	102400	409600
14	6 µg FimCH	25600	409600	409600	204800	819200
15	1.2 µg FimCH	100	25600	25600	25600	204800
16	1.2 µg FimCH	100	12800	6400	6400	102400

Example 5

Low Dose dscFimH

Further experiments were performed to show that acceptable responses could be achieved at lower doses. Immunogenicity was determined (using MF59 as adjuvant) at several lower doses. The results are presented in Tables 4, 5 and 6. In general, even at a dose as low as 0.32 µg (following a boost) or 2 µg (pre-boost) good responses were observed. For this experiment, C3H/HeJ mice were immunized on day 0

and boosted at week 4 (5 mice/group). EPTs (end point titers) represent titers for pooled sera for each group. Immun responses were measured pre- and post-boost against three different capture antigens (by ELISA): FimH T3 (a FimH truncate), dscFimH and FimH (0.4 M Urea).

Table 4

Immunogen (dose)	α-FimH T3 (3 weeks)	α-FimH T3 (8 weeks)
FimHdsc (2 μg)	1600	102400
FimHdsc (0.8 μg)	<100	51200
FimHdsc (0.32 μg)	<100	12800
FimCH (4 μg)	51200	204800
FimCH (0.26 μg)	6400	409600

Table 5

Immunogen (dose)	α-FimH T3 (3 weeks)	α-FimHdsc (8 weeks)
FimHdsc (2 μg)	6400	204800
FimHdsc (0.8 μg)	200	204800
FimHdsc (0.32 μg)	<100	51200
FimCH (4 μg)	51200	409600
FimCH (0.26 μg)	6400	409600

Table 6

Immunogen (dose)	α-FimH T3 (3 wks)	α-FimH (.4M Urea) (8 wks)
FimHdsc (2 µg)	800	51200
FimHdsc (0.8 µg)	<100	102400
FimHdsc (0.32 µg)	<100	12800
FimCH (4 µg)	25600	204800
FimCH (0.26 µg)	1600	102400

Claims

WHAT IS CLAIMED IS:

1. An immunogenic complex comprising a pilus-protein component and a donor strand component.

2. The immunogenic complex of claim 1 wherein said pilus-protein component is covalently bound to said donor strand component.

3. The immunogenic complex of claim 1 wherein said pilus-protein component is non-covalently bound to said donor strand component.

4. The immunogenic complex of claim 1 wherein said pilus protein is an adhesin.

5. The immunogenic complex of claim 4 wherein said adhesin is FimH.

6. The immunogenic complex of claim 1 wherein the donor strand is selected from the group consisting of SEQ ID NOs: 3, 4, 5, 6, 7, and 8.

7. A polypeptide comprising a pilus-protein portion and a donor complement portion as part of the same amino acid sequence.

8. The polypeptide of claim 7 wherein the donor complement portion is derived from a bacterial chaperone.

9. The polypeptide of claim 7 wherein the donor complement portion is a bacterial pilin or adhesin, a portion thereof, or derived from a bacterial pilin or adhesin.

10. The polypeptide of claim 8 wherein the chaperone is selected from the group consisting of PapD and FimC.

11. The polypeptide of claim 8 wherein the bacterial chaperone is derived from *E. coli*.

12. The polypeptide of claim 7 wherein the donor complement portion is selected from the group consisting of SEQ ID NOS: 3, 4, 5, 6, 7, and 8.

13. The polypeptide of claim 7 wherein the pilus-protein portion is found in bacteria of the family *enterobacteriaceae*.

14. The polypeptide of claim 13 wherein the bacterium is *E. coli*.

15. The polypeptide of claim 7 wherein the pilus-protein is an adhesin.

16. The polypeptide of claim 7 wherein the pilus-protein is a pilin.

17. The polypeptide of claim 7 wherein the pilus protein is PapK.

18. The polypeptide of claim 15 wherein the adhesin is FimH.

19. The polypeptide of claim 18 having a sequence selected from the group consisting of SEQ ID NO: 10 and 11.

20. A single polypeptide or polypeptide complex comprising the polypeptide of claim 7 and an adhesin.

21. A single polypeptide or polypeptide complex comprising the polypeptide of claim 7 and a pilin.

22. The polypeptide complex of claim 20 and 21 wherein the pilus-protein portion of the polypeptide is FimG and the adhesin is FimH.

23. A polynucleotide comprising a coding region for the polypeptide of claim 7, 20 or 21.

24. An antibody specific for an immunogenic complex selected from the group consisting of the polypeptides and/or complexes of claims 1, 2, 3, 4, 5, and 6.

25. An antibody specific f r a polypeptide selected from the group consisting of the polypeptides of claims 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, and 22.

26. The antibody of claim 24 or 25 wherein said antibody is a monoclonal antibody.

27. A genetically engineered cell expressing the antibody of claim 26.

28. A genetically engineered cell expressing the polynucleotide of claim 23.

29. A vector comprising the polynucleotide of claim 23.

30. A genetically engineered cell expressing the polypeptide of claim 19.

31. A composition comprising the antibody of claim 24 or 25 suspended in a pharmacologically acceptable carrier, diluent or excipient.

32. A composition comprising an immunogenic complex selected from the group consisting of the complexes of claims 1, 2, 3, 4, 5, and 6, said immunogenic complex being suspended in a pharmacologically acceptable carrier, diluent or excipient.

33. A composition comprising a polypeptide selected from the group consisting of the polypeptides of claims 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, and 22, said polypeptide being suspended in a pharmacologically acceptable carrier, diluent or excipient.

34. A vaccine composition comprising an immunogenically effective amount of an immunogenic complex selected from the group consisting of the complexes of claims 1, 2, 3, 4, 5, and 6, said immunogenic complex being suspended in a pharmacologically acceptable carrier, diluent or

excipient.

35. A vaccine composition comprising an immunogenically effective amount of a polypeptide selected from the group consisting of the polypeptides of claims 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, and 22, said polypeptide being suspended in a pharmacologically acceptable carrier, diluent or excipient.

36. A method of preventing a disease in a mammal at risk thereof comprising administering to said animal the vaccine composition of claim 34 or 35.

37. The method of claim 36 wherein the disease is caused by a bacterium of the family *enterobacteriaceae*.

38. The method of claim 37 wherein the bacterium is *E. coli*.

39. The method of claim 30 wherein the mammal is a human.

40. The method of claim 36 wherein the disease is a urinary tract disease.

41. A method of treating a disease in a mammal afflicted therewith comprising administering to said animal a pharmacologically effective amount of the composition of claim 31.

42. The method of claim 41 wherein the disease is caused by a bacterium of the family *enterobacteriaceae*.

43. The method of claim 42 wherein the bacterium is *E. coli*.

44. The method of claim 41 wherein the animal is a human.

45. The method of claim 44 wherein the disease is a urinary tract

infection.

46. The method of claim 41 wherein the disease is caused by a bacterium of the family *Enterobacteriaceae.*

47. The method of claim 46 wherein the disease is caused by *E. coli.*

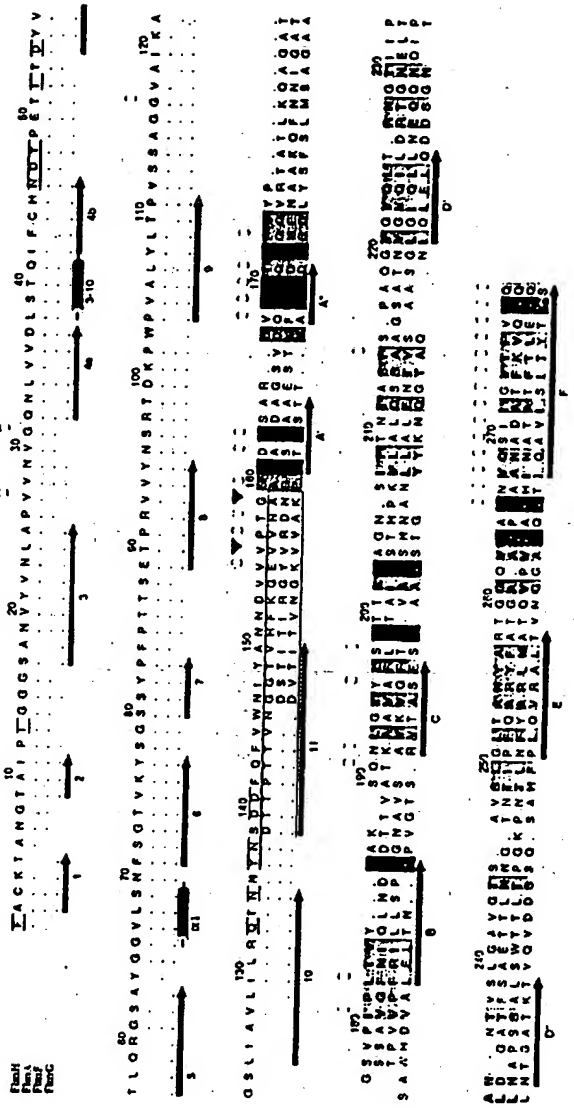
48. A method of producing an immunogenic polypeptide comprising synthesizing a pilus-polypeptide having a donor complement strand attached thereto.

49. The method of claim 48 wherein the donor complement strand is selected from the group consisting of SEQ ID NOS: 3, 4, 5, 6, 7, and 8.

50. The method of claim 48 wherein the pilus-polypeptide is an adhesin.

51. The method of claim 50 wherein the adhesin is FimH.

Fig. 1



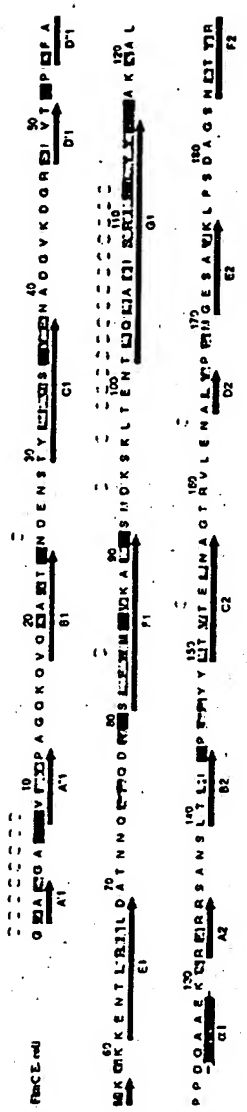


Fig. 2

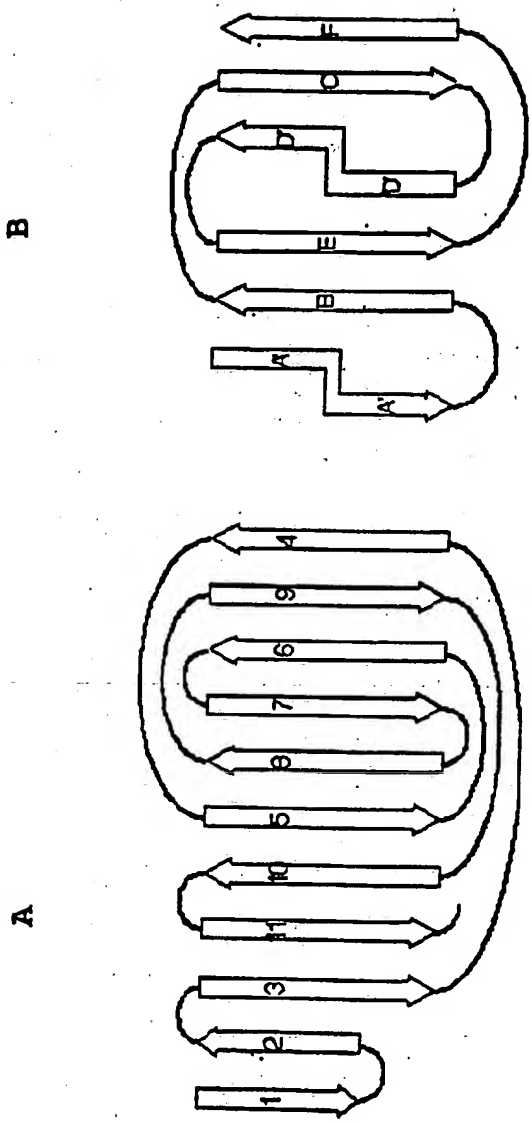


Fig. 3

Fig. 4

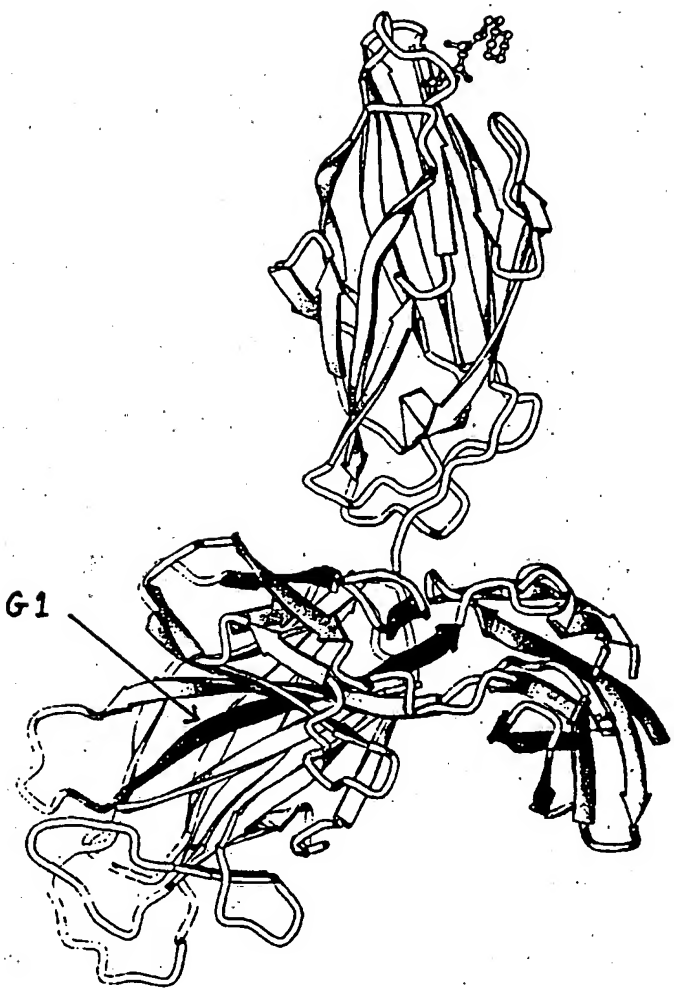


Fig 5

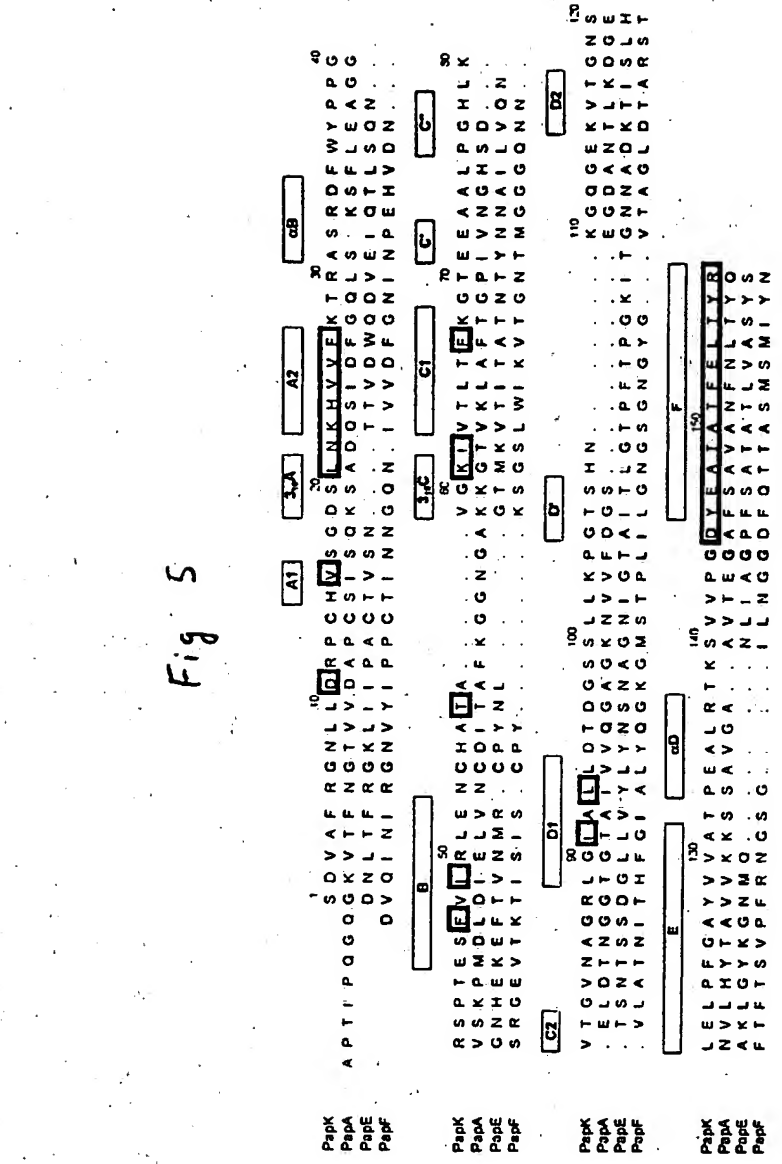
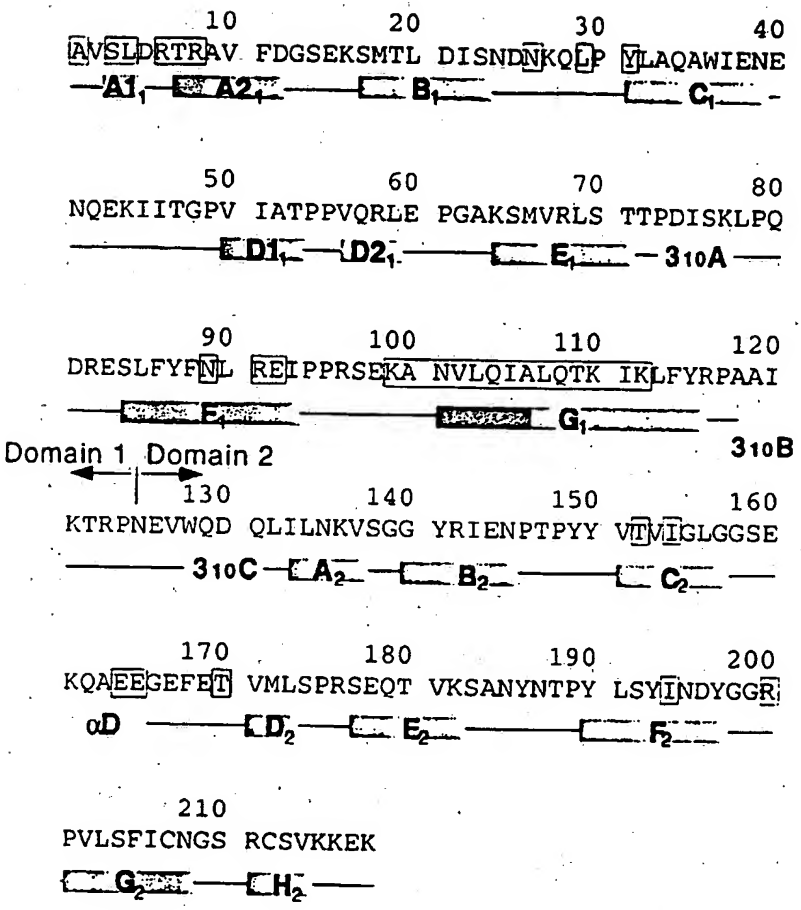


Fig. 6


10 20 30 40
AVSLDRTRAV FDGSEKSMTL DISNDNKQLP VLAQAWIENE
A1 A2 B C
50 60 70 80
NQEKIITGPV IATPPVQRLE PGAKSMVRLS TTPDISKLPQ
D1 D2 E 310A
90 100 110 120
DRESLFYFNL REIPPRSEKA NVLQIALQTK IKLFYRPAAI
F G
Domain 1 Domain 2
310B
130 140 150 160
KTRPNEVWQD QLILNKVSGG YRIENPTPYY VTVIGLGGSE
310C A2 B2 C2
170 180 190 200
KQAEEGEFET VMLSPRSEQT VKSANYNTPY LSYINDYGGR
αD D2 E2 F2
210
PVLSFICNGS RCSVKKEK
G2 H2

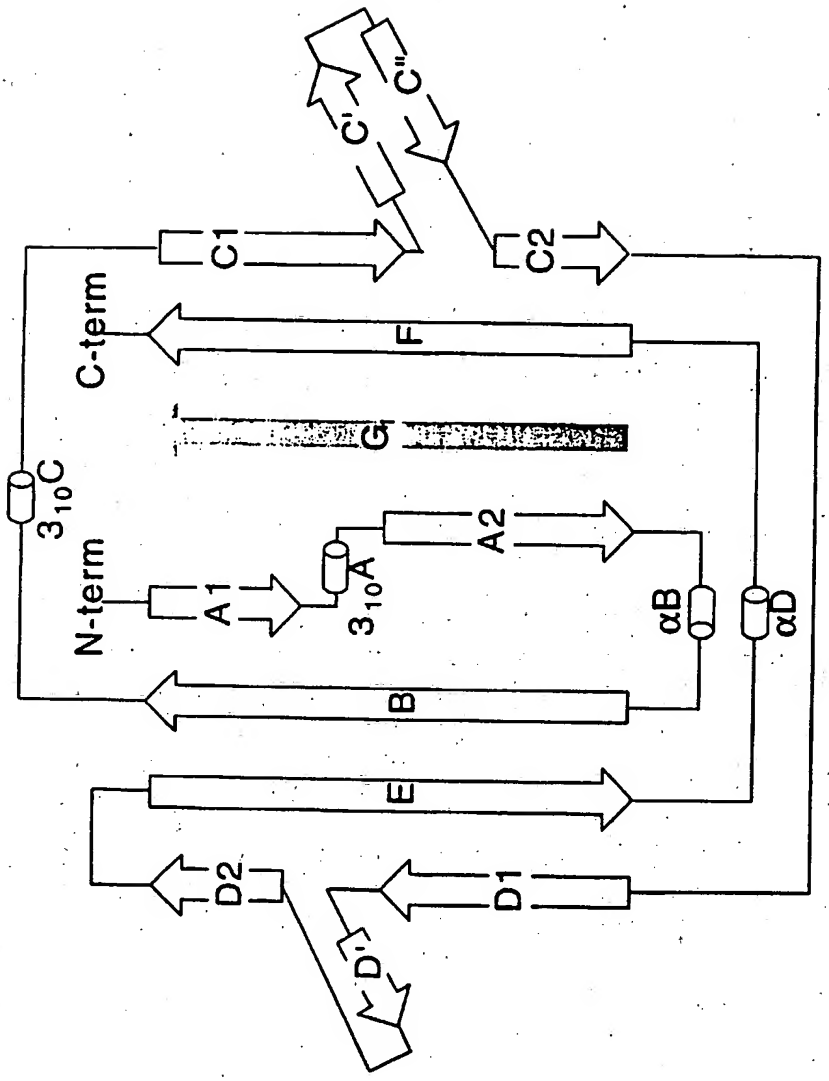


Fig. 7

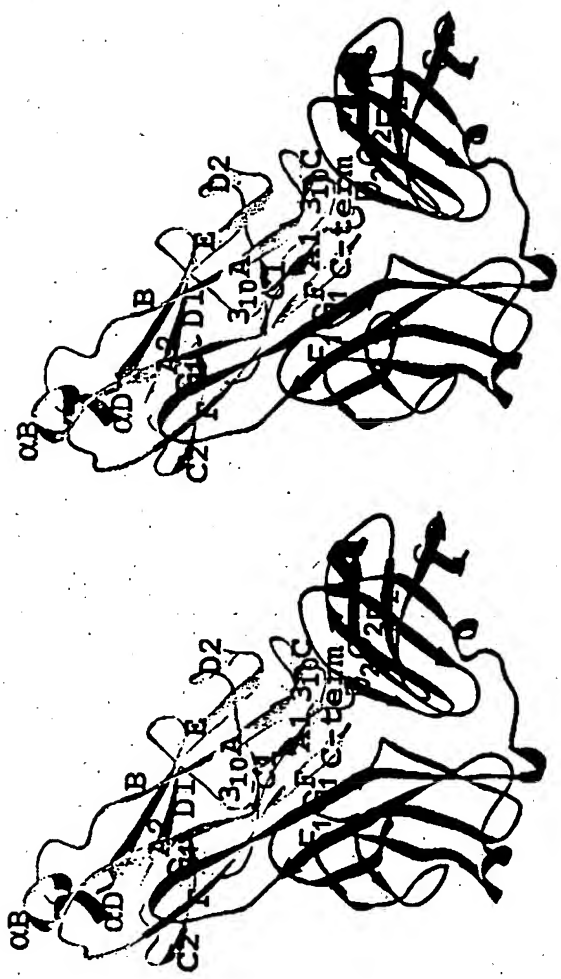


Fig. 8

Fig. 9

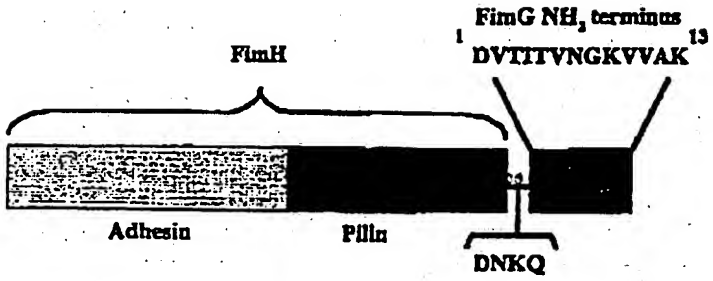
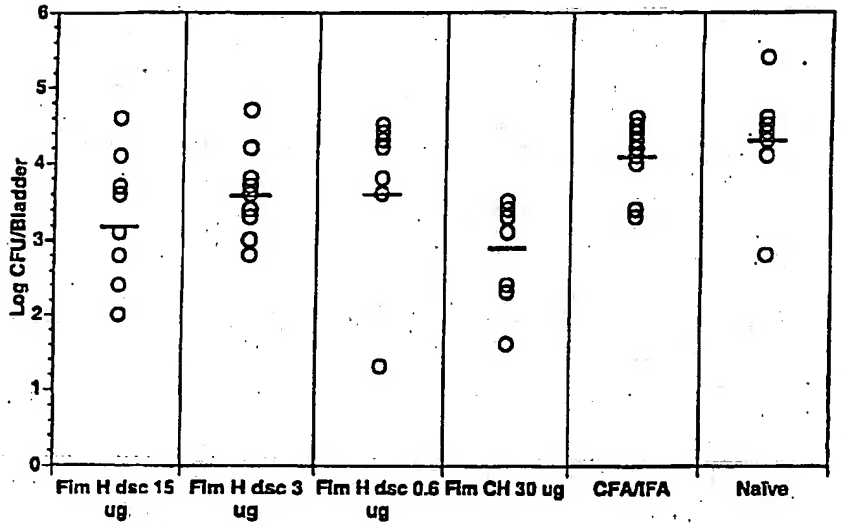

FimG NH2 terminus
1
DVTITVNGKVVAK
13
FimH
Adhesin
Pilin
DNKQ

Fig. 10

C3H/HeJ mice immunized with Pilus Protein + CFA/IFA
4 wk boost IM 9 wk Challenge IU with 11.3 x 10^7 cfu/dose NU 14



T Test	CFA/IFA	Naive
15 ug Fim H donor strand complement	0.02	0.006
3 ug Fim H donor strand complement	0.05	0.02
0.6 ug Fim H donor strand complement	0.288	0.117
30 ug Fim CH	0.0001	0.0001

Fig. 11

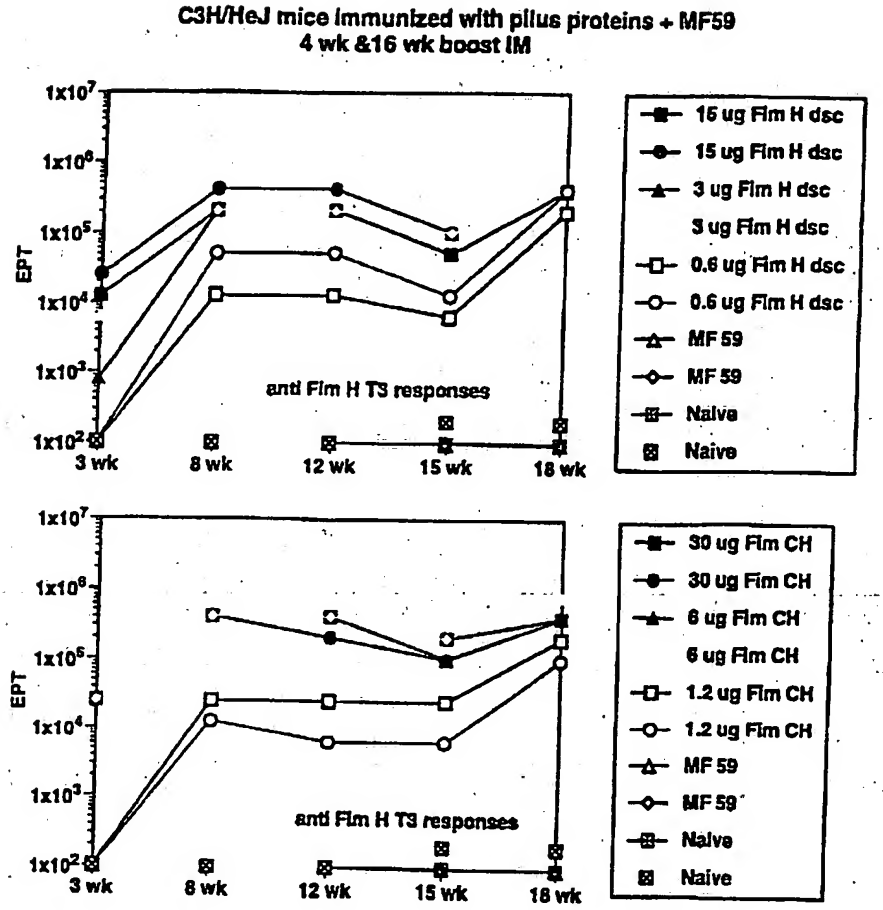
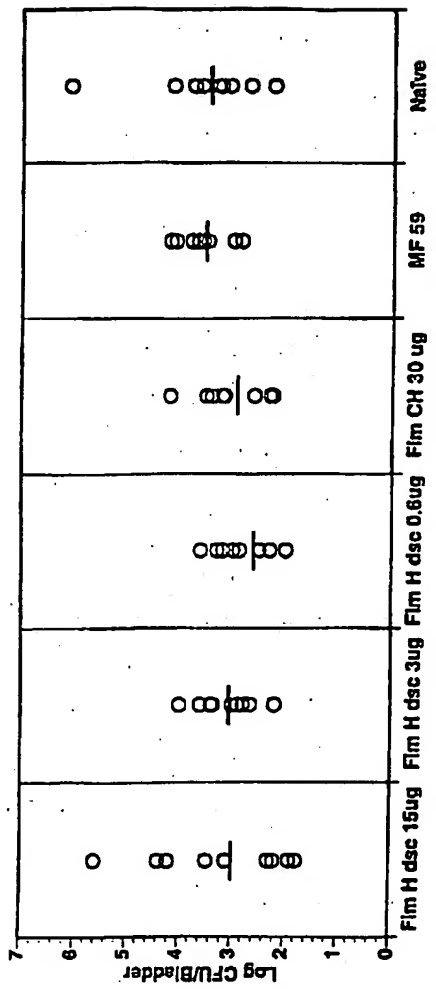

C3H/HeJ mice immunized with pilus proteins + MF59
4 wk &16 wk boost IM
EPT
1x10^7
1x10^6
1x10^5
1x10^4
1x10^3
1x10^2
anti Fim H T3 responses
3 wk
8 wk
12 wk
15 wk
18 wk
15 ug Fim H dsc
15 ug Fim H dsc
3 ug Fim H dsc
3 ug Fim H dsc
0.6 ug Fim H dsc
0.6 ug Fim H dsc
MF 59
MF 59
Naive
Naive
EPT
1x10^7
1x10^6
1x10^5
1x10^4
1x10^3
1x10^2
anti Fim H T3 responses
3 wk
8 wk
12 wk
15 wk
18 wk
30 ug Fim CH
30 ug Fim CH
6 ug Fim CH
6 ug Fim CH
1.2 ug Fim CH
1.2 ug Fim CH
MF 59
MF 59
Naive
Naive

Fig. 12



T TEST	MF 59	Naïve
15 ug Flm H DSC	0.45	0.53
3 ug Flm H dsc	0.07	0.28
0.6 ug Flm H dsc	0.001	0.03
30 ug Flm CH	0.047	0.189

SEQUENCE LISTING

<110> Hultgren, Scott J.
Pinkner, Jerome S.
Sauer, Frederic
Barnhart, Michelle
Waksman, Gabriel
Knight, Stefan

<120> Donor Strand Complemented Pilin and Adhesin Broad-Based Vaccines

<130> 469201-479

<140>
<141>

<150> U.S. 60/143,582
<151> 1999-07-13

<150> U.S. 60/144,359
<151> 1999-07-16

<150> U.S. 60/184,442
<151> 2000-02-23

<160> 23

<170> PatentIn Ver. 2.1

<210> 1
<211> 279
<212> PRT
<213> Escherichia coli

<400> 1

```
Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
 1               5                  10                  15

Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val Asn Val Gly Gln
            20                  25                  30

Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
        35                  40                  45

Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
    50                  55                  60
```

```
Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly Ser
 65                  70                  75                  80

Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
                85                  90                  95

Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
            100                 105                 110

Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
        115                 120                 125

Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
    130                 135                 140

Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
145                 150                 155                 160

Cys Asp Val Ser Ala Arg Asn Val Thr Val Thr Leu Pro Asp Tyr Pro
                165                 170                 175

Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
            180                 185                 190

Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly Asn Ser Ile
        195                 200                 205

Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
    210                 215                 220

Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
225                 230                 235                 240

Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
                245                 250                 255

Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
            260                 265                 270

Gly Val Thr Phe Val Tyr Gln
        275
```

<210> 2
<211> 205
<212> PRT
<213> Escherichia coli

<400> 2

Gly Val Ala Leu Gly Ala Thr Arg Val Ile Tyr Pro Ala Gly Gln Lys
1 5 10 15

Gln Val Gln Leu Ala Val Thr Asn Asn Asp Glu Asn Ser Thr Tyr Leu
20 25 30

Ile Gln Ser Trp Val Glu Asn Ala Asp Gly Val Lys Asp Gly Arg Phe
35 40 45

Ile Val Thr Pro Pro Leu Phe Ala Met Lys Gly Lys Lys Glu Asn Thr
50 55 60

Leu Arg Ile Leu Asp Ala Thr Asn Asn Gln Leu Pro Gln Asp Arg Glu
65 70 75 80

Ser Leu Phe Trp Met Asn Val Lys Ala Ile Pro Ser Met Asp Lys Ser
85 90 95

Lys Leu Thr Glu Asn Thr Leu Gln Leu Ala Ile Ile Ser Arg Ile Lys
100 105 110

Leu Tyr Tyr Arg Pro Ala Lys Leu Ala Leu Pro Pro Asp Gln Ala Ala
115 120 125

Glu Lys Leu Arg Phe Arg Arg Ser Ala Asn Ser Leu Thr Leu Ile Asn
130 135 140

Pro Thr Pro Tyr Tyr Leu Thr Val Thr Glu Leu Asn Ala Gly Thr Arg
145 150 155 160

Val Leu Glu Asn Ala Leu Val Pro Pro Met Gly Glu Ser Ala Val Lys
165 170 175

Leu Pro Ser Asp Ala Gly Ser Asn Ile Thr Tyr Arg Thr Ile Asn Asp
180 185 190

Tyr Gly Ala Leu Thr Pro Lys Met Thr Gly Val Met Glu
195 200 205

<210> 3
<211> 15
<212> PRT
<213> Escherichia coli

<400> 3
Thr Leu Gln Leu Ala Ile Ile Ser Arg Ile Lys Leu Tyr Tyr Arg

1 5 10 15

<210> 4
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Reverse of the G1 strand of FimC shown in SEQ ID NO: 3.

<400> 4
Arg Tyr Tyr Leu Lys Ile Arg Ser Ile Ile Ala Leu Gln Leu Thr
1 5 10 15

<210> 5
<211> 14
<212> PRT
<213> Escherichia coli

<400> 5
Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Pro
1 5 10

<210> 6
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Reverse of the PapG donor strand shown in SEQ ID NO: 5.

<400> 6
Pro Lys Ala Val Val Lys Gly Asn Val Thr Ile Thr Val Asp
1 5 10

<210> 7
<211> 13
<212> PRT
<213> Escherichia coli

<400> 7
Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys

1 5 10

<210> 8
<211> 13
<212> PRT
<213> Escherichia coli

<400> 8
Asp Ser Thr Ile Thr Ile Arg Gly Tyr Val Arg Asp Asn
1 5 10

<210> 9
<211> 143
<212> PRT
<213> Escherichia coli

<400> 9
Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Pro Cys Thr
1 5 10 15

Val Ser Thr Thr Asn Ala Thr Val Asp Leu Gly Asp Leu Tyr Ser Phe
20 25 30

Ser Leu Met Ser Ala Gly Ala Ala Ser Ala Trp His Asp Val Ala Leu
35 40 45

Glu Leu Thr Asn Cys Pro Val Gly Thr Ser Arg Val Thr Ala Ser Phe
50 55 60

Ser Gly Ala Ala Asp Ser Thr Gly Tyr Tyr Lys Asn Gln Gly Thr Ala
65 70 75 80

Gln Asn Ile Gln Leu Glu Leu Gln Asp Asp Ser Gly Asn Thr Leu Asn
85 90 95

Thr Gly Ala Thr Lys Thr Val Gln Val Asp Asp Ser Ser Gln Ser Ala
100 105 110

His Phe Pro Leu Gln Val Arg Ala Leu Thr Val Asn Gly Gly Ala Thr
115 120 125

Gln Gly Thr Ile Gln Ala Val Ile Ser Ile Thr Tyr Thr Tyr Ser
130 135 140

<210> 10

<211> 304
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:DscFimH with G1 strand of FimC and linker

<400> 10

```
Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
  1               5                  10                  15

Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val Asn Val Gly Gln
             20                  25                  30

Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
         35                  40                  45

Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
     50                  55                  60

Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly Ser
 65                  70                  75                  80

Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
                 85                  90                  95

Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
            100                 105                 110

Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
        115                 120                 125

Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
    130                 135                 140

Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
145                 150                 155                 160

Cys Asp Val Ser Ala Arg Asn Val Thr Val Thr Leu Pro Asp Tyr Pro
                165                 170                 175

Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
            180                 185                 190

Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly Asn Ser Ile
        195                 200                 205
```

```
Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
    210                 215                 220

Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
225                 230                 235                 240

Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
                245                 250                 255

Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
            260                 265                 270

Gly Val Thr Phe Val Tyr Gln Gly Ser Gly Ser Gly Ser Gly Ser Gly
        275                 280                 285

Ser Thr Leu Gln Leu Ala Ile Ile Ser Arg Ile Lys Leu Tyr Tyr Arg
    290                 295                 300
```

<210> 11
<211> 296
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:DscFimH with donor strand from PimG and linker

<400> 11

```
Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
  1               5                  10                  15

Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val Asn Val Gly Gln
            20                  25                  30

Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
        35                  40                  45

Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
    50                  55                  60

Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly Ser
65                  70                  75                  80

Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
```

```
                85                  90                  95

Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
            100                 105                 110

Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
        115                 120                 125

Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
    130                 135                 140

Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
145                 150                 155                 160

Cys Asp Val Ser Ala Arg Asn Val Thr Val Thr Leu Pro Asp Tyr Pro
                165                 170                 175

Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
            180                 185                 190

Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly Asn Ser Ile
        195                 200                 205

Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
    210                 215                 220

Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
225                 230                 235                 240

Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
                245                 250                 255

Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
            260                 265                 270

Gly Val Thr Phe Val Tyr Gln Asp Asn Lys Gln Asp Val Thr Ile Thr
        275                 280                 285

Val Asn Gly Lys Val Val Ala Lys
    290                 295


<210> 12
<211> 153
<212> PRT
<213> Escherichia coli

<400> 12
```

```
Asp Ser Thr Ile Thr Ile Arg Gly Tyr Val Val Asn Ala Ala Cys Ala
 1               5                  10                  15

Val Asp Ala Gly Ser Val Asp Gln Thr Val Gln Leu Gly Gln Val Arg
            20                  25                  30

Thr Ala Thr Leu Lys Gln Ala Gly Ala Thr Ser Ser Ala Val Gly Phe
        35                  40                  45

Asn Ile Gln Leu Asn Asp Cys Asp Thr Thr Val Ala Thr Lys Ala Ala
    50                  55                  60

Val Ala Phe Leu Gly Thr Ala Ile Asp Ser Thr His Pro Lys Val Leu
65                  70                  75                  80

Ala Leu Gln Ser Ser Ala Ala Gly Ser Ala Thr Asn Val Gly Val Gln
                85                  90                  95

Ile Leu Asp Arg Thr Gly Asn Glu Leu Thr Leu Asp Gly Ala Thr Phe
            100                 105                 110

Ser Ala Glu Thr Thr Leu Asn Asn Gly Thr Asn Thr Ile Pro Phe Gln
        115                 120                 125

Ala Arg Tyr Phe Ala Thr Gly Ala Ala Thr Pro Gly Ala Ala Asn Ala
    130                 135                 140

Asp Ala Thr Phe Lys Val Gln Tyr Gln
145                 150
```

<210> 13
<211> 152
<212> PRT
<213> Escherichia coli

<400> 13

```
Asp Ser Thr Ile Thr Ile Arg Gly Tyr Val Arg Asp Asn Gly Cys Ser
 1               5                  10                  15

Val Ala Ala Glu Ser Thr Asn Phe Thr Val Asp Leu Met Glu Asn Ala
            20                  25                  30

Ala Lys Gln Phe Asn Asn Ile Gly Ala Thr Thr Pro Val Val Pro Phe
        35                  40                  45

Arg Ile Leu Leu Ser Pro Cys Gly Asn Ala Val Ser Ala Val Lys Val
    50                  55                  60
```

```
Gly Phe Leu Gly Val Ala Asp Ser His Asn Ala Asn Leu Leu Ala Leu
 65                  70                  75                  80

Glu Asn Thr Val Ser Ala Ala Ser Gly Leu Gly Ile Gln Leu Leu Asn
                 85                  90                  95

Glu Gln Asn Gln Ile Pro Leu Asn Ala Pro Ser Ser Ala Leu Ser Trp
            100                 105                 110

Thr Thr Leu Thr Pro Gly Lys Pro Asn Thr Leu Asn Phe Tyr Ala Arg
        115                 120                 125

Leu Met Ala Thr Gln Val Pro Val Thr Ala Gly His Ile Asn Ala Thr
    130                 135                 140

Ala Thr Phe Thr Leu Glu Tyr Gln
145                 150


<210> 14
<211> 218
<212> PRT
<213> Escherichia coli

<400> 14
Ala Val Ser Leu Asp Arg Thr Arg Ala Val Phe Asp Gly Ser Glu Lys
  1                 5                  10                  15

Ser Met Thr Leu Asp Ile Ser Asn Asp Asn Lys Gln Leu Pro Tyr Leu
             20                  25                  30

Ala Gln Ala Trp Ile Glu Asn Glu Asn Gln Glu Lys Ile Ile Thr Gly
         35                  40                  45

Pro Val Ile Ala Thr Pro Pro Val Gln Arg Leu Glu Pro Gly Ala Lys
     50                  55                  60

Ser Met Val Arg Leu Ser Thr Thr Pro Asp Ile Ser Lys Leu Pro Gln
 65                  70                  75                  80

Asp Arg Glu Ser Leu Phe Tyr Phe Asn Leu Arg Glu Ile Pro Pro Arg
                 85                  90                  95

Ser Glu Lys Ala Asn Val Leu Gln Ile Ala Leu Gln Thr Lys Ile Lys
            100                 105                 110

Leu Phe Tyr Arg Pro Ala Ala Ile Lys Thr Arg Pro Asn Glu Val Trp
                        10
```

```
        115                 120                 125

Gln Asp Gln Leu Ile Leu Asn Lys Val Ser Gly Gly Tyr Arg Ile Glu
    130                 135                 140

Asn Pro Thr Pro Tyr Tyr Val Thr Val Ile Gly Leu Gly Gly Ser Glu
145                 150                 155                 160

Lys Gln Ala Glu Glu Gly Glu Phe Glu Thr Val Met Leu Ser Pro Arg
                165                 170                 175

Ser Glu Gln Thr Val Lys Ser Ala Asn Tyr Asn Thr Pro Tyr Leu Ser
            180                 185                 190

Tyr Ile Asn Asp Tyr Gly Gly Arg Pro Val Leu Ser Phe Ile Cys Asn
        195                 200                 205

Gly Ser Arg Cys Ser Val Lys Lys Glu Lys
    210                 215


<210> 15
<211> 157
<212> PRT
<213> Escherichia coli

<400> 15
Ser Asp Val Ala Phe Arg Gly Asn Leu Leu Asp Arg Pro Cys His Val
  1               5                  10                  15

Ser Gly Asp Ser Leu Asn Lys His Val Val Phe Lys Thr Arg Ala Ser
            20                  25                  30

Arg Asp Phe Trp Tyr Pro Pro Gly Arg Ser Pro Thr Glu Ser Phe Val
        35                  40                  45

Ile Arg Leu Glu Asn Cys His Ala Thr Ala Val Gly Lys Ile Val Thr
    50                  55                  60

Leu Thr Phe Lys Gly Thr Glu Glu Ala Ala Leu Pro Gly His Leu Lys
65                  70                  75                  80

Val Thr Gly Val Asn Ala Gly Arg Leu Gly Ile Ala Leu Leu Asp Thr
                85                  90                  95

Asp Gly Ser Ser Leu Leu Lys Pro Gly Thr Ser His Asn Lys Gly Gln
            100                 105                 110
```

Gly Glu Lys Val Thr Gly Asn Ser Leu Glu Leu Pro Phe Gly Ala Tyr
115 120 125

Val Val Ala Thr Pro Glu Ala Leu Arg Thr Lys Ser Val Val Pro Gly
130 135 140

Asp Tyr Glu Ala Thr Ala Thr Phe Glu Leu Thr Tyr Arg
145 150 155

<210> 16
<211> 163
<212> PRT
<213> Escherichia coli

<400> 16
Ala Pro Thr Ile Pro Gln Gly Gln Gly Lys Val Thr Phe Asn Gly Thr
1 5 10 15

Val Val Asp Ala Pro Cys Ser Ile Ser Gln Lys Ser Ala Asp Gln Ser
20 25 30

Ile Asp Phe Gly Gln Leu Ser Lys Ser Phe Leu Glu Ala Gly Gly Val
35 40 45

Ser Lys Pro Met Asp Leu Asp Ile Glu Leu Val Asn Cys Asp Ile Thr
50 55 60

Ala Phe Lys Gly Gly Asn Gly Ala Lys Lys Gly Thr Val Lys Leu Ala
65 70 75 80

Phe Thr Gly Pro Ile Val Asn Gly His Ser Asp Glu Leu Asp Thr Asn
85 90 95

Gly Gly Thr Gly Thr Ala Ile Val Val Gln Gly Ala Gly Lys Asn Val
100 105 110

Val Phe Asp Gly Ser Glu Gly Asp Ala Asn Thr Leu Lys Asp Gly Glu
115 120 125

Asn Val Leu His Tyr Thr Ala Val Val Lys Lys Ser Ser Ala Val Gly
130 135 140

Ala Ala Val Thr Glu Gly Ala Phe Ser Ala Val Ala Asn Phe Asn Leu
145 150 155 160

Thr Tyr Gln

<210> 17
<211> 148
<212> PRT
<213> Escherichia coli

<400> 17
Asp Asn Leu Thr Phe Arg Gly Lys Leu Ile Ile Pro Ala Cys Thr Val
1 5 10 15

Ser Asn Thr Thr Val Asp Trp Gln Asp Val Glu Ile Gln Thr Leu Ser
20 25 30

Gln Asn Gly Asn His Glu Lys Glu Phe Thr Val Asn Met Arg Cys Pro
35 40 45

Tyr Asn Leu Gly Thr Met Lys Val Thr Ile Thr Ala Thr Asn Thr Tyr
50 55 60

Asn Asn Ala Ile Leu Val Gln Asn Thr Ser Asn Thr Ser Ser Asp Gly
65 70 75 80

Leu Leu Val Tyr Leu Tyr Asn Ser Asn Ala Gly Asn Ile Gly Thr Ala
85 90 95

Ile Thr Leu Gly Thr Pro Phe Thr Pro Gly Lys Ile Thr Gly Asn Asn
100 105 110

Ala Asp Lys Thr Ile Ser Leu His Ala Lys Leu Gly Tyr Lys Gly Asn
115 120 125

Met Gln Asn Leu Ile Ala Gly Pro Phe Ser Ala Thr Ala Thr Leu Val
130 135 140

Ala Ser Tyr Ser
145

<210> 18
<211> 148
<212> PRT
<213> Escherichia coli

<400> 18
Asp Val Gln Ile Asn Ile Arg Gly Asn Val Tyr Ile Pro Pro Cys Thr
1 5 10 15

```
Ile Asn Asn Gly Gln Asn Ile Val Val Asp Phe Gly Asn Ile Asn Pro
            20                  25                  30

Glu His Val Asp Asn Ser Arg Gly Glu Val Thr Lys Thr Ile Ser Ile
        35                  40                  45

Ser Cys Pro Tyr Lys Ser Gly Ser Leu Trp Ile Lys Val Thr Gly Asn
    50                  55                  60

Thr Met Gly Gly Gly Gln Asn Asn Val Leu Ala Thr Asn Ile Thr His
65                  70                  75                  80

Phe Gly Ile Ala Leu Tyr Gln Gly Lys Gly Met Ser Thr Pro Leu Ile
                85                  90                  95

Leu Gly Asn Gly Ser Gly Asn Gly Tyr Gly Val Thr Ala Gly Leu Asp
            100                 105                 110

Thr Ala Arg Ser Thr Phe Thr Phe Thr Ser Val Pro Phe Arg Asn Gly
        115                 120                 125

Ser Gly Ile Leu Asn Gly Gly Asp Phe Gln Thr Thr Ala Ser Met Ser
    130                 135                 140

Met Ile Tyr Asn
145
```

<210> 19
<211> 337
<212> PRT
<213> Escherichia coli

<400> 19

```
Met Lys Lys Trp Phe Pro Ala Leu Leu Phe Ser Leu Cys Val Ser Gly
  1               5                  10                  15

Glu Ser Ser Ala Trp Asn His Asn Ile Val Phe Tyr Ser Leu Gly Asn
            20                  25                  30

Val Asn Ser Tyr Gln Gly Gly Asn Val Val Ile Thr Gln Arg Pro Gln
        35                  40                  45

Phe Ile Thr Ser Trp Arg Pro Gly Ile Ala Thr Val Thr Trp Asn Gln
    50                  55                  60

Cys Asn Gly Pro Glu Phe Ala Asp Gly Ser Trp Ala Tyr Tyr Arg Glu
65                  70                  75                  80
```

```
Tyr Ile Ala Trp Val Val Phe Pro Lys Lys Val Met Thr Gln Asn Gly
                85                  90                  95

Tyr Pro Leu Phe Ile Glu Val His Asn Lys Gly Ser Trp Ser Glu Glu
           100                 105                 110

Asn Thr Gly Asp Asn Asp Ser Tyr Phe Phe Leu Lys Gly Tyr Lys Trp
       115                 120                 125

Asp Glu Arg Ala Phe Asp Ala Gly Asn Leu Cys Gln Lys Pro Gly Glu
   130                 135                 140

Thr Thr Arg Leu Thr Glu Lys Phe Asp Asp Ile Ile Phe Lys Val Ala
145                 150                 155                 160

Leu Pro Ala Asp Leu Pro Leu Gly Asp Tyr Ser Val Thr Ile Pro Tyr
               165                 170                 175

Thr Ser Gly Ile Gln Arg His Phe Ala Ser Tyr Leu Gly Ala Arg Phe
           180                 185                 190

Lys Ile Pro Tyr Asn Val Ala Lys Thr Leu Pro Arg Glu Asn Glu Met
       195                 200                 205

Leu Phe Leu Phe Lys Asn Ile Gly Gly Cys Arg Pro Ser Ala Gln Ser
   210                 215                 220

Leu Glu Ile Lys His Gly Asp Leu Ser Ile Asn Ser Ala Asn Asn His
225                 230                 235                 240

Tyr Ala Ala Gln Thr Leu Ser Val Ser Cys Asp Val Pro Ala Asn Ile
               245                 250                 255

Arg Phe Met Leu Leu Arg Asn Thr Thr Pro Thr Tyr Ser His Gly Lys
           260                 265                 270

Lys Phe Ser Val Gly Leu Gly His Gly Trp Asp Ser Ile Val Ser Val
       275                 280                 285

Asn Gly Val Asp Thr Gly Glu Thr Thr Met Arg Trp Tyr Lys Ala Gly
   290                 295                 300

Thr Gln Asn Leu Thr Ile Gly Ser Arg Leu Tyr Gly Glu Ser Ser Lys
305                 310                 315                 320

Ile Gln Pro Gly Val Leu Ser Gly Ser Ala Thr Leu Leu Met Ile Leu
               325                 330                 335
```

Pro

<210> 20
<211> 13
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Donor strand derived from the first 13 residues of PapF

<400> 20
Asp Val Gln Ile Asn Ile Arg Gly Asn Val Tyr Ile Pro
1 5 10

<210> 21
<211> 354
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Donor strand complemented form of PapG (SEQ ID NO:19)

<400> 21
Met Lys Lys Trp Phe Pro Ala Leu Leu Phe Ser Leu Cys Val Ser Gly
1 5 10 15

Glu Ser Ser Ala Trp Asn His Asn Ile Val Phe Tyr Ser Leu Gly Asn
20 25 30

Val Asn Ser Tyr Gln Gly Gly Asn Val Val Ile Thr Gln Arg Pro Gln
35 40 45

Phe Ile Thr Ser Trp Arg Pro Gly Ile Ala Thr Val Thr Trp Asn Gln
50 55 60

Cys Asn Gly Pro Glu Phe Ala Asp Gly Ser Trp Ala Tyr Tyr Arg Glu
65 70 75 80

Tyr Ile Ala Trp Val Val Phe Pro Lys Lys Val Met Thr Gln Asn Gly
85 90 95

Tyr Pro Leu Phe Ile Glu Val His Asn Lys Gly Ser Trp Ser Glu Glu

100 105 110

Asn Thr Gly Asp Asn Asp Ser Tyr Phe Phe Leu Lys Gly Tyr Lys Trp
115 120 125

Asp Glu Arg Ala Phe Asp Ala Gly Asn Leu Cys Gln Lys Pro Gly Glu
130 135 140

Thr Thr Arg Leu Thr Glu Lys Phe Asp Asp Ile Ile Phe Lys Val Ala
145 150 155 160

Leu Pro Ala Asp Leu Pro Leu Gly Asp Tyr Ser Val Thr Ile Pro Tyr
165 170 175

Thr Ser Gly Ile Gln Arg His Phe Ala Ser Tyr Leu Gly Ala Arg Phe
180 185 190

Lys Ile Pro Tyr Asn Val Ala Lys Thr Leu Pro Arg Glu Asn Glu Met
195 200 205

Leu Phe Leu Phe Lys Asn Ile Gly Gly Cys Arg Pro Ser Ala Gln Ser
210 215 220

Leu Glu Ile Lys His Gly Asp Leu Ser Ile Asn Ser Ala Asn Asn His
225 230 235 240

Tyr Ala Ala Gln Thr Leu Ser Val Ser Cys Asp Val Pro Ala Asn Ile
245 250 255

Arg Phe Met Leu Leu Arg Asn Thr Thr Pro Thr Tyr Ser His Gly Lys
260 265 270

Lys Phe Ser Val Gly Leu Gly His Gly Trp Asp Ser Ile Val Ser Val
275 280 285

Asn Gly Val Asp Thr Gly Glu Thr Thr Met Arg Trp Tyr Lys Ala Gly
290 295 300

Thr Gln Asn Leu Thr Ile Gly Ser Arg Leu Tyr Gly Glu Ser Ser Lys
305 310 315 320

Ile Gln Pro Gly Val Leu Ser Gly Ser Ala Thr Leu Leu Met Ile Leu
325 330 335

Pro Asp Asn Lys Gln Asp Val Gln Ile Asn Ile Arg Gly Asn Val Tyr
340 345 350

Ile Pro

<210> 22
<211> 84
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Oligonucleotide
used in cloning

<400> 22
cgattattgg cgtgactttt gtttatcaag ataacaaaca ggatgtcacc atcacggtga 60
acggtaaggt cgtcgccaaa taag 84

<210> 23
<211> 87
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Oligonucleotide
used in cloning

<400> 23
gatccttatt tggcgacgac cttaccgttc accgtgatgg tgaccatcct gtttgttatc 60
ttgataaaca aaagtcacgc caataat 87